

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	84.9	100.0	3837	1	STACNA		MB1736 StephyLococ
2	84.9	100.0	3837	6	AR067706		AR067706 Sequence
3	792.2	93.3	3468	6	AI18434		AI18434 Hybrid DNA
4	792.2	93.3	4612	6	AI18436		AI18436 assembled s
5	792.2	93.3	4612	6	AR067712		AR067712 Sequence
6	73.4	8.6	155204	2	AC007926		AC007926 Trypanoso
7	65.6	7.7	171798	2	AC084397		AC084397 Trypanoso
8	64.2	7.6	7218	6	I66494		I66494 Sequence 14
9	63.2	7.4	155204	2	AC007926		AC007926 Trypanoso
10	62.8	7.4	208546	2	AC083827		AC083827 Homo sapi
11	61.8	7.3	1822	1	AF260878		AF260878 Enterococ
12	61.8	7.3	318231	2	PFMAL13P3		AL049184 Plasmodiu
13	61.4	7.2	104992	2	AC005504		AC005504 Plasmodiu
14	61.4	7.2	169546	2	AC004157		AC004157 Plasmodiu
15	60.6	7.1	110000	2	PFMAL41_1		Continuation (2 of
16	60.2	7.1	957	1	AF260880		AF260880 Enterococ
17	60.2	7.1	957	1	AF260881		AF260881 Enterococ
18	60.2	7.1	957	1	AF260882		AF260882 Enterococ
19	60.2	7.1	957	1	AF260883		AF260883 Enterococ
20	60.2	7.1	957	1	AF260888		AF260888 Enterococ
21	60.2	7.1	957	1	AF260890		AF260890 Enterococ
22	60.2	7.1	957	1	AF260893		AF260893 Enterococ
23	60.2	7.1	957	1	AF260895		AF260895 Enterococ
24	60.2	7.1	1963	1	AF260876		AF260876 Enterococ
25	58.8	6.9	192929	2	AC005505		AC005505 Plasmodiu
26	58.8	6.9	256172	2	AC005139		AC005139 Plasmodiu
27	58.6	6.9	863	11	CNS06EVO		ALJ95628 T7 end o
28	58.6	6.9	957	1	AF260889		AF260889 Enterococ
29	58.6	6.9	957	1	AF260894		AF260894 Enterococ
30	58.6	6.9	1963	1	AF260877		AF260877 Enterococ
31	58.6	6.9	2104	1	AF260874		AF260874 Enterococ
32	58.6	6.9	2104	1	AF260875		AF260875 Enterococ
33	58.6	6.9	2245	1	AF260872		AF260872 Enterococ
34	58.6	6.9	2245	1	AF260873		AF260873 Enterococ
35	58	6.8	3989	14	EPVSPIER		X77052 Entomopoxv
36	57.6	6.8	110000	2	PFMAL41_0		AL034357 Plasmodiu
37	57.4	6.8	4096	3	AF362374		AF362374 Dicycsto
38	57.4	6.8	47577	3	AF396446		AF396446 Tetrahyme
39	57.4	6.8	189883	2	AC013254		AC013254 Drosophil
40	57	6.7	957	1	AF260884		AF260884 Enterococ
41	57	6.7	957	1	AF260885		AF260885 Enterococ
42	57	6.7	957	1	AF260886		AF260886 Enterococ
43	57	6.7	957	1	AF260892		AF260892 Enterococ
44	57	6.7	957	1	AF260896		AF260896 Enterococ
45	57	6.7	1374	1	AF159247		AF159247 Enterococ

RESULT

STACNA	LOCUS	3827 bp	DNA	BCF	17-AUG-1994
DEFINITION	Staphylococcus aureus	collagen adhesin (cna) gene, complete cds			
ACCESSION	M81736				
VERSION	M81736.1	GI:387879			
KEYWORDS	cna gene; collagen adhesin.				
SOURCE	Staphylococcus aureus DNA.				
ORGANISM	Staphylococcus aureus				
	Bacteria; Firmicutes: Bacillus/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus.				
REFERENCE	1 (bases 1 to 3827)				
AUTHORS	Pati, J.M., Jonsson, H., Guss, B., Switalski, L.M., Wilberg, K.,				

TITLE	Molecular characterization and expression of a gene encoding a Staphylococcus aureus collagen adhesin [published sequence appears in J Biol Chem 1994 Apr 15; 269(15):11672]
JOURNAL	J. Biol. Chem. 267 (7), 4766-4772 (1992)
MEDLINE	92165839
COMMENT	On Aug 30, 1993 this sequence version replaced g1:386712.

FEATURES
Source
RBS
gene
CDS

Location/Qualifiers
1..3827
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
/issue_11b="FDA 574; lambda gtl1"
137..144
/gene="cna"
137..3702
/gene="cna"
151..3702
/gene="cna"
/codon_start=1
/product="collagen adhesin"
/protein_id="AA020874.1"
/db_xref="GI:387880"
/translation="MNKLVKMFVIMLNIITPLFNKNEAFANDISITVITLTVS
PSKIDGGKTVKTFKIDKNGKIDGMDKIVAMPISGVKIGSKYVPLVYKSGV
QAVITPDGATITFENDKVERKIDSGFAEVEVGRLTQNTSDKVAITTSKNTY
TVKSEAGTSVFFYKTDMLPEDTTHRMPLININNESYSKDTIKDQIGGOLD
LSTLAINITGHSNYSOSAITDEKAFGSKITVDNTKNTIDVTIDQIGSIVSFS
INVKITITNEQKEFNNSQAWYQEHKFEVNGKSFNHNINAGIEGVKELK
VAKODRKAPIANVKFKSKDGSVKNQKEIITDANGIANIKALPSGDYLIKE
IEAPRYFDKDEKPEFMKIDPDNOGYFTTINAKAIEKTKVSAOKWEGIOYKPT
IYFKLYKODDNGNTVPDKAEIKLEDGKTKVTGSLNPENDKNGKAIYIYEVNAG
EDTPEGYTKKNGLVNTNTEKPIETTSISGEKWDKDNODKREKYSVNLANGE
KKTITDVTSETMKTEFKDLPKIDEGKILEYVTDHVKDYDTIDNGTITNKYPGE
TSATVTKMWDNNKDGKRPTEIKVELYODGATGKTALINESNMWHTMGLDEKAK
GOOVYVVELTKVGGYTHVDNMDGNLIVNKYTPETTSISGEKWDKDNODKRE
PEKVSYNLADDEKVKITLDVSETNMKYEKDPYDEKILEYVTEHVDYDTID
NGTITNKYTPETSATVTKMWDNNKDGKRPTEIKVELYODGATGKTALINESN
WHTMGLDEKAKGOOVYVVELTKVGGYTHVDNMDGNLIVNKYTPETTSISGE
KWDKDNODKREKYSVNLANGEKVTLDVSETNMKYEKDPYDEKILEYV
VTDHVKDYDTIDNGTITNKYTPETSATVTKMWDNNKDGKRPTEIKVELYODGK
ATGKTAALINESNMWHTMGLDEKAKGOOVYVVELTKVNGYTHVDNMDGNLIV
NTEKPKPKPIYEPKPKDPKDPKDPKDPKDPKDPKDPKDPKDPKDPKDPKDPK
PKELPKTKMKTITSMITWVFIGIIGIYILKRRNS"

BASE COUNT 1668 a 533 c 737 g 889 t
ORIGIN

Query Match 100.0%; Score 849; DB 1; Length 3827;
Best Local Similarity 100.0%; Pred. No. 9.3e-137;
Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCATAAAAATGAAAAATATCAAAATGTGACATGATTAAGTGGCATGCCGACACAGC 60
|||||
DB 331 GAGCATAAAAATGAAAAATATCAAAATGTGACATGATTAAGTGGCATGCCGACACAGC 330
|||||
QY 61 GGTACAGTAAAGATAGAGGTTTACTAAAAACGTACCATTAACCTGTTAAAGGTGAACAG 120
|||||
DB 391 GGTACAGTAAAGATAGAGGTTTACTAAAAACGTACCATTAACCTGTTAAAGGTGAACAG 450
|||||
QY 121 GTGGGTCAACGAGTTTATACACAGAGGTGCAACAATTCAATTCATGATTAAGTAAAGTAA 180
|||||
DB 451 GTGGGTCAACGAGTTTATACACAGAGGTGCAACAATTCAATTCATGATTAAGTAAAGTAA 510
|||||
QY 181 AAATTAAGTATGTTTCGGATTTGCAGAATTTGAAGTACAGAGAAATTTTAAACGCA 240
|||||
DB 511 AAATTAAGTATGTTTCGGATTTGCAGAATTTGAAGTACAGAGAAATTTTAAACGCA 570
|||||
QY 241 ACAAAATCTTCAGATGACAAAGTACATTAACATCTGGGAATTAATCAACGAATGTT 300
|||||
DB 571 ACAAAATCTTCAGATGACAAAGTACATTAACATCTGGGAATTAATCAACGAATGTT 630
|||||
QY 301 ACGGTTCAATAAAGTGAAGGGGGAACAAGTATGTTTCTATATATAAAGGGAGATATG 360
|||||
DB 631 ACGGTTCAATAAAGTGAAGGGGGAACAAGTATGTTTCTATATATAAAGGGAGATATG 690
|||||
QY 361 CTACACAGAAATACGACATGATGATGTTTAAATTAATTAACATGAAAAAGTTAT 420
|||||
DB 691 CTACACAGAAATACGACATGATGATGTTTAAATTAATTAACATGAAAAAGTTAT 750
|||||

QY 421 GTATCGAAAAGATATTACTATTAAGATGATTCAGATTCAGGTGGACACAGTATGATTTAAGC 480
|||||
DB 751 GTATCGAAAAGATATTACTATTAAGATGATTCAGATTCAGGTGGACACAGTATGATTTAAGC 810
|||||
QY 481 ACATTAACATTAATGTGACAGTACACATACATCAATTTATATAGTGACAAAGTCAATT 540
|||||
DB 811 ACATTAACATTAATGTGACAGTACACATACATCAATTTATATAGTGACAAAGTCAATT 870
|||||
QY 541 ACTGATTTTGAAGAAAGCCTTTCAGAGTTCTTAAATTAAGTGTGATTAATACGAAGACACA 600
|||||
DB 871 ACTGATTTTGAAGAAAGCCTTTCAGAGTTCTTAAATTAAGTGTGATTAATACGAAGACACA 930
|||||
QY 601 ATTGATGTAAACATTCACACAGGCTATGCGTATTAATAGTTTTCATTAACATCAACAA 660
|||||
DB 931 ATTGATGTAAACATTCACACAGGCTATGCGTATTAATAGTTTTCATTAACATCAACAA 990
|||||
QY 661 ACCAAATATACAAATGAACAGCAAAAAGAGTTGTTAATTAATTCACAGCTGTATCA 720
|||||
DB 991 ACCAAATATACAAATGAACAGCAAAAAGAGTTGTTAATTAATTCACAGCTGTATCA 1050
|||||
QY 721 GAGCATGTAAAGAAAGTGAACGGGAAATCATTTAATCATCTGTGCACAAATTAAT 780
|||||
DB 1051 GAGCATGTAAAGAAAGTGAACGGGAAATCATTTAATCATCTGTGCACAAATTAAT 1110
|||||
QY 781 GCTAATGCCGATTTGCAAGCTACTGTAAGAAAGTGAATTAAGTTTAAACAGATTA 840
|||||
DB 1111 GCTAATGCCGATTTGCAAGCTACTGTAAGAAAGTGAATTAAGTTTAAACAGATTA 1170
|||||
QY 841 GATACCAAG 849
|||||
DB 1171 GATACCAAG 1179
|||||

RESULT 2
LOCUS AR067706 3827 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5851794.
ACCESSION AR067706
VERSION AR067706.1 GI:5998928
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3827)
AUTHORS Guss,B., Hook,M., Jonsson,H., Lindberg,M., Patil,D., Signas,C. and
Switala,S.L.
TITLE Collagen binding protein as well as its preparation
JOURNAL Patent: US 5851794-A 1 22-DEC-1998;
FEATURES
source location/Qualifiers
1..3827
/organism="unknown"
BASE COUNT 1668 a 533 c 737 g 889 t
ORIGIN

Query Match 100.0%; Score 849; DB 6; Length 3827;
Best Local Similarity 100.0%; Pred. No. 9.3e-137;
Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCATAAAAATGAAAAATATCAAAATGTGACATGATTAAGTGGCATGCCGACACAGC 60
|||||
DB 331 GAGCATAAAAATGAAAAATATCAAAATGTGACATGATTAAGTGGCATGCCGACACAGC 390
|||||
QY 61 GGTACAGTAAAGATAGAGGTTTACTAAAAACGTACCATTAACCTGTTAAAGGTGAACAG 120
|||||
DB 391 GGTACAGTAAAGATAGAGGTTTACTAAAAACGTACCATTAACCTGTTAAAGGTGAACAG 450
|||||
QY 121 GTGGGTCAACGAGTTTATACACAGAGGTGCAACAATTCAATTCATGATTAAGTAAAGTAA 180
|||||
DB 451 GTGGGTCAACGAGTTTATACACAGAGGTGCAACAATTCAATTCATGATTAAGTAAAGTAA 510
|||||
QY 181 AAATTAAGTATGTTTCGGATTTGCAGAATTTGAAGTACAGAGAAATTTTAAACGCA 240
|||||

Db	511	AAATTAAGTATGTTTGGCATTTGCAGAAATTTGAGTACAAAGGAAGAAATTTACGCA	570
Qy	241	ACAAATFACCTCAGATGACAAAGTACGTACGATTAACATCTGGGAATAAATCAACGATGTT	300
Db	571	ACAAATFACCTCAGATGACAAAGTACGTACGATTAACATCTGGGAATAAATCAACGATGTT	630
Qy	301	ACGGTTCATTAAGTGAAGCGGGAACAGTAGTGTTTTCTATTAAACCGGACATATG	360
Db	631	ACGGTTCATTAAGTGAAGCGGGAACAGTAGTGTTTTCTATTAAACCGGACATATG	690
Qy	361	CTACCCAGAGATACGACACATGTACGATGTTTTTAAATATTAACATGAAAAAGTTAT	420
Db	691	CTACCCAGAGATACGACACATGTACGATGTTTTTAAATATTAACATGAAAAAGTTAT	750
Qy	421	GTATCGAAAGATATTACTATTAAGGATCAGATTCAGGTGACAGCTAGATTAAAGC	480
Db	751	GTATCGAAAGATATTACTATTAAGGATCAGATTCAGGTGACAGCTAGATTAAAGC	810
Qy	481	ACATTAACCATTAATGTACAGGTACATATGCAATTTATTATGTGACAAAGTGCATTT	540
Db	811	ACATTAACCATTAATGTACAGGTACATATGCAATTTATTATGTGACAAAGTGCATTT	870
Qy	541	ACTGATTTTGAAGAAAGCTTTCCAGGTTCTAAATATACGTGTGATTAATACGAGAACA	600
Db	871	ACTGATTTTGAAGAAAGCTTTCCAGGTTCTAAATATACGTGTGATTAATACGAGAACA	930
Qy	601	ATTGATGTACAAATTTCCAAAGCTATGCGTCATTAATAGTTTTTCAATTAACACAAA	660
Db	931	ATTGATGTACAAATTTCCAAAGCTATGCGTCATTAATAGTTTTTCAATTAACACAAA	990
Qy	661	ACCAAAATTTAGCAATGACACGAAAGAAAGTTGTTAATTAATTCACAGCTGGATCAA	720
Db	991	ACCAAAATTTAGCAATGACACGAAAGAAAGTTGTTAATTAATTCACAGCTGGATCAA	1050
Qy	721	GAGCATGTGAAGAAAGAGTGAACGGGAATCATTTATCATCTACGTGCACAAATTTAAT	780
Db	1051	GAGCATGTGAAGAAAGAGTGAACGGGAATCATTTATCATCTACGTGCACAAATTTAAT	1110
Qy	781	GCTAATGCCGCTATTGAAGTACTGTAAAGGTCAATTTAAAGGTTTAAACAGGATAAA	840
Db	1111	GCTAATGCCGCTATTGAAGTACTGTAAAGGTCAATTTAAAGGTTTAAACAGGATAAA	1170
Qy	841	GATACCAAG 849	
Db	1171	GATACCAAG 1179	
RESULT	3		
LOCUS	A18434	3468 bp DNA	PAT 18-Apr-1994
DEFINITION	A18434	Hybrid DNA molecule comprising a nucleotide sequence from S.aureus coding for a protein or polypeptide having collagen binding activity.	
ACCESSION	A18434		
VERSION	A18434.1	GI:513301	
KEYWORDS			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
SOURCE			
CDS			

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
93.3%; Score 792.2; DB 6; Length 3468;	97.8%; Pred. No. 5.3e-127;	836;	0;	13;	6;	3;
Query	1	GACGATTAATAATGGAAAAATACAAAATGTCACATGATTAAGAATGCGATGGCCGCAACG	60			
Db	94	GACGATTAATAATGGAAAAATACAAAATGTCACATGATTAAGAATGCGATGGCCGCAACG	153			
Query	61	GGTACAGTAAGAATGAGAGGTTATGTAATAACAGATTAACCTGTTAAAGAGTAACAG	120			
Db	154	GGTACAGTAAGAATGAGAGGTTATGTAATAACAGATTAACCTGTTAAAGAGTAACAG	213			
Query	121	GTGGTCAAGCAGTTATTACACACAGCGGTGCACAAATTAATCAATGATTAAGATAGAA	180			
Db	214	GTGGTCAAGCAGTTATTACACACAGCGGTGCACAAATTAATCAATGATTAAGATAGAA	273			
Query	181	AAATTAAGTATGTTCTGGCGATTTGCACAATTTGAAGTACACAGGAAGAAATTTAACGAA	240			
Db	274	AAATTAAGTATGTTCTGGCGATTTGCACAATTTGAAGTACACAGGAAGAAATTTAACGAA	333			
Query	241	ACCAATTAATTCAGATGACAAAGTACTCGATACATCTGGGAATTAATCAACGAATGTT	300			
Db	334	ACCAATTAATTCAGATGACAAAGTACTCGATACATCTGGGAATTAATCAACGAATGTT	393			
Query	301	A - -CGGTCATTAATAAGTGAACGCGGGAACAAGTAGTGTTCATATATATAA - - -CGGAG	355			
Db	394	ATCGGTGATTAATAAGTGAACGCGGGAACAAGTAGTGTTCATATATATAAAGCGGGAAG	453			
Query	356	ATATGCTAC - AGAAGATACGACACATGTACAGAGGTTTTPAATATTPAACATGAAGAA	414			
Db	454	ATATGCTACCAAGAAAGATACACACATGTACAGAGGTTTTPAATATTPAACATGAAGAA	513			
Query	415	AGTTATGATTAAGAAAGATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT	474			
Db	514	AGTTATGATTAAGAAAGATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT	573			
Query	475	TTAAGCAGATTAATAAGTGAACGAGTACATTAACCAATTAATTAAGTGAACAAAGT	534			
Db	574	TTAAGCAGATTAATAAGTGAACGAGTACATTAACCAATTAATTAAGTGAACAAAGT	633			
Query	535	GCAATTAAGTATTTGAAGAAAGCTTTCAGAGTTCATAAATTAAGTGTGATTAATGCAAG	594			
Db	634	GCAATTAAGTATTTGAAGAAAGCTTTCAGAGTTCATAAATTAAGTGTGATTAATGCAAG	693			
Query	595	AACCAATTAAGTATTAAGCAATTCACAAAGGCTATAGGCTATTAATTAAGTGTTCATTAAC	654			

```

Db 694 AACACATTGATGTAACCAATTCACCAAGGCTATGGGTCAATATAAGTTTTCATTAAC 753
QY 655 TACAAACCAAAATTAAGTAAGTAAGCAAGAAAGAGTTGTTTAATTAATTCACAAAGCTGG 714
Db 754 TACAAACCAAAATTAAGTAAGTAAGCAAGAAAGAGTTGTTTAATTAATTCACAAAGCTGG 813
QY 715 TATCAAGAGCATGTGTAAGGAAGTAAGGGAATTCATTTAATCATCTGTGCACAT 774
Db 814 TATCAAGAGCATGTGTAAGGAAGTAAGGGAATTCATTTAATCATCTGTGCACAT 873
QY 775 ATTAATGCTAATCCGGTATGTAAGGTACTGTAAAGGTATTAAGTTTAAACAG 834
Db 874 ATTAATGCTAATCCGGTATGTAAGGTACTGTAAAGGTATTAAGTTTAAACAG 933
QY 835 GATTAAGATACCAAG 849
Db 934 GATTAAGATACCAAG 948

RESULT 4
LOCUS A18436 4612 bp DNA PAT 18-APR-1994
DEFINITION assembled sequence from the insert in p 16 and cCOLR6a.
ACCESSION A18436
VERSION A18436.1 GI:513303
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4612)
AUTHORS
TITLE A COLLAGEN BINDING PROTEIN AS WELL AS ITS PREPARATION
JOURNAL Patent: WO 9207002-A 3 30-APR-1992;
FEATURES
source
1. 4612
/organism="synthetic construct"
/db_xref="taxon:32630"
/codon_start=1
/transl_table=11
/product="polypeptide with collagen binding activity"
/protein_id="CA01396.1"
/db_xref="GI:513304"
/translation="MKNVNLKFMVFTMLNITPLEFNKNEAPARDISTNVTLTFS
PSKIDEGKTYVMTFDDKNGKILONGDMKIVAMPTSGTVKIEGSKTVPLTVKGEQV
QAVITPDGATITFENDKVEKLSVSGFAEEVQGNLTOTNTLDDKVAITTSNKSSTNV
IGMIVKREPVEFLINKSGKICGYOEDTTHVRMLNNKSKSYSKDITTKDIOGGQO
LILSTLINIVTGHSHSYSGASATIDFEKAPGSKTYNTNTIDVTIPDGYGSYS
ESINKTKRTTNQOKREFVNSQATVOEHKEKEVNGKSEPHYTHININAGISTYKGE
LAVLQODKTPKAPLANVKFKLSKDSGVKDNOKETIITDANGIANIKALPSGDYIL
KEIEARPYTFEDKDEYPTMDTDNOGFTTIEANKALEKRDVSAQVWEGTOVK
PIYFELVYKODNONMTTPVDKAEIKKLEGGTKVWSNLPENDKNGKAIKYVKEVNA
OSEDTPPEGYTKKENGIVYTNTEKPIETTSISEGKWDMDKNDKDRPRKPSVNLAN
GKSVKTLDTSETNKKIEFKDLPKYDEGKIEYTVEDHVDKDYTTDINGTTTNTKYP
GERTSATVKNMDNNODGKRPETKVELYQGGKATGKATILNESNHTHTTGTDEK
AKGOQKYTVBELTKVKGTTTVDDNDMGNLIVTNKYTPETTSISEGKWDMDKNDG
KRPERSVNLADGKVKTLDTVSETNMYEKERDLPKYDEGKIEYTVEDHVDKDYTT
DINGTTTNTKYPGETSATVKNMDNNODGKRPETKVELYQGGKATGKATILNES
NMWHTWTGLDEKAKGOQKYTVBELTKVKGTTTVDDNDMGNLIVTNKYTPETTSIS
GKRWMDKNDGKDRPRKPSVNLANGEVKTLDVTSNMYEKERDLPKYDEGKIE
YTVEDHVDKDYTTDINGTTTNTKYPGETSATVKNMDNNODGKRPETKVELYQD
GKATGKTALINESNMWHTWTGLDEKAKGOQKYTVDELTKVNGYTHADNDMGNLI
VNNKYTPKRPKPIYEPKPKDPTPTPKPSHKSQVKTPTPDKPSKVDKDPDKNTKP
ENPLKELPTGKMKIITSWITWFIIGILYLIRKRFS"
BASE COUNT 1946 a 676 c 868 g 1122 t
ORIGIN

```

```

Query Match 93.3%; Score 792.2; DB 6; Length 4612;
Best Local Similarity 97.8%; Pred. No. 5,1e-127;
Matches 836; Conservative 0; Mismatches 13; Indels 6; Gaps 3;
QY 1 GACGATTAATAATGAAAAATACAAATGTGACATGATTAAGTGGCAGCCGACAAAGC 60

```

```

Db 1111 GACGATTAATAATGAAAAATACAAATGTGACATGATTAAGTGGCAGCCGACAAAGC 1170
QY 61 GGTACAGTAAGATAGAGGTTTATAGTAAACAGTACCAATTAAGTGTAAAGGTGAACAG 120
Db 1171 GGTACAGTAAGATAGAGGTTTATAGTAAACAGTACCAATTAAGTGTAAAGGTGAACAG 1230
QY 121 GTGGGTCAAGCAGTATTAACACGACGCGTGCACAAATTAATCAATTAAGTGAAGAA 180
Db 1231 GTGGGTCAAGCAGTATTAACACGACGCGTGCACAAATTAATCAATTAAGTGAAGAA 1290
QY 181 AATTAAGTATGCTTTCGGGATTTGAGAAATTAAGTCAAGGAAATTAACGCCAA 240
Db 1291 AATTAAGTATGCTTTCGGGATTTGAGAAATTAAGTCAAGGAAATTAACGCCAA 1350
QY 241 ACAATTAATCTTACAGTACGACAAAGTAGTACGATTAATCAATCTGGGAATTAACGAATGTT 300
Db 1351 ACAATTAATCTTACAGTACGACAAAGTAGTACGATTAATCAATCTGGGAATTAACGAATGTT 1410
QY 301 A--CGGTTCAATAAAGTGAAGCGGGAACAAGTAGTGTCTTAATTAATAA--CGGGAG 355
Db 1411 ATCGGTTGATTAAGTGAAGCGGGAACAAGTAGTGTCTTAATTAATAAAGCGGGAAG 1470
QY 356 ATATGCTACC-AGAAGATACGACACATGTCATGCTTTTAAATATTAACATGAAGAA 414
Db 1471 ATATGCTACCAGAAAGATACGACACATGTCATGCTTTTAAATATTAACATGAAGAA 1530
QY 415 AGTTATGATTCGAAGATTTACTATAAAGGATCAATTAAGGATGACAGCAGTATGAT 474
Db 1531 AGTTATGATTCGAAGATTTACTATAAAGGATCAATTAAGGATGACAGCAGTATGAT 1590
QY 475 TTAAAGCATTAATAACATTAATGTCAGAGGTACACATAGCAATTAATTAAGTGAACAAAGT 534
Db 1591 TTAAAGCATTAATAACATTAATGTCAGAGGTACACATAGCAATTAATTAAGTGAACAAAGT 1650
QY 535 GCATTAATGATTTTGAAGAAAGCCTTCCAGGTTCTTAATAATCACTGTGATTAACGAAG 594
Db 1651 GCATTAATGATTTTGAAGAAAGCCTTCCAGGTTCTTAATAATCACTGTGATTAACGAAG 1710
QY 595 AACCAATGATGATTAACCAATTCACCAAGGCTATGGGTCAATTAATGATTTTCAATTAAC 654
Db 1711 AACCAATGATGATTAACCAATTCACCAAGGCTATGGGTCAATTAATGATTTTCAATTAAC 1770
QY 655 TACAAACCAAAATTAAGTAAGTAAGCAAGCAAGAAAGAGTTGTTAATTAATTCACAAAGCTGG 714
Db 1771 TACAAACCAAAATTAAGTAAGTAAGCAAGCAAGAAAGAGTTGTTAATTAATTCACAAAGCTGG 1830
QY 715 TATCAAGAGCATGTGTAAGGAAGTAAGGGAATTCATTTAATCATCTGTGCACAT 774
Db 1831 TATCAAGAGCATGTGTAAGGAAGTAAGGGAATTCATTTAATCATCTGTGCACAT 1890
QY 775 ATTAATGCTAATCCGGTATGTAAGGTACTGTAAAGGTATTAAGTTTAAACAG 834
Db 1891 ATTAATGCTAATCCGGTATGTAAGGTACTGTAAAGGTATTAAGTTTAAACAG 1950
QY 835 GATTAAGATACCAAG 849
Db 1951 GATTAAGATACCAAG 1965

RESULT 5
LOCUS AR067712 4612 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 8 from patent US 5851794.
ACCESSION AR067712
VERSION AR067712.1 GI:5998934
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4612)
AUTHORS Guss,B., Hook,M., Jonsson,H., Lindberg,M., Patil,J., Signas,C. and Switalski,I.

```


[illegible]

```

RESULT      7
AC084397/c
LOCUS       AC084397      171798 bp    DNA
DEFINITION  Trypanosoma brucei chromosome unknown clone RCI193-1118. ***
ACCESSION   AC084397
VERSION     AC084397.4  GI:11612622
KEYWORDS    HTGS_PHASE2.
SOURCE      Trypanosoma brucei.
ORGANISM    Trypanosoma brucei

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 171798)
El-Sayed, N.M., Ghedin, E., Song, J., Larkin, C., Mantless, D., Jones, K.,
Peterson, J., Hou, L., Zhao, H., Mason, T., Miltseth, J., Pal, G., Van
Aken, S., Utterback, T., Khalak, H.G., Gerard, C., Leech, V., Jilly, E.,
Melville, S., White, O., Adams, M.D., Donelson, J.E. and Fraser, C.M.
Trypanosoma brucei GUTt10.1 RFL193-III8 BAC genomic sequence
Unpublished
2 (bases 1 to 171798)
El-Sayed, N.M., Khalak, H. and Adams, M.D.
Direct Submission
Submitted (30-OCT-2000) The Institute for Genomic Research, 9712

ON DEC 11, 2000 this sequence version replaced 011123084.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 171798: contig of 171798 bp in length.

```

FEATURES
  source
    1..171798
      /organism="Trypanosoma brucei"
      /isolate="GUT410.1"
      /db_xref="taxon:5691"
      /chromosome="unknown"
      /clone="Rc193-1118"
BASE COUNT
  39792 a 37575 c 33263 g 61168 t
ORIGIN

```

Query Match 7.7% Score 65.6 DB 2 Length 171798;
Best Local Similarity 46.8% Pred. No. 0.016;
Matches 206; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY	457	GGTGGACGCGAGTTAGATTTTAAAGACATTTAAACCTTATGTGGACAGGTACACATTCAGAT	516
Db	158664	AAT	158605
QY	517	TATTATAGTGACAAAGTGCATTTACTGATTTTGGAAAAACCTTCCAGGTCTAAAAATA	576
Db	158604	AAT	158545
QY	577	ACTGTGTATATACGAAGACACACATTGATGTAAACAATTCACAGGCTATGGGTCATAT	636
Db	158544	AAT	158485
QY	637	AATAGTTTTCAATTAACACCAAAACCAAAATTCGAATGCACGCAAAAGAGTTGGT	696
Db	158444	AAT	158425
QY	697	AATATATCCAAAGCTGGTATCAAGACATGTGAAGAGAGTGAACGGAAATCATTT	756
Db	158424	AAT	158365
QY	757	AATCACTACTGTGCACAAATATTATGCTAATGCGGATATTCAGGACTGTAAAAAGTGAA	816
Db	158364	AAT	158305
QY	817	TTAAAGTTTTAAAAACAGA	836
Db	158304	GTACTAATATTATAGTATGA	158285

RESULT		8	PAT	28-DEC-1997
LOCUS	166494/C			
DEFINITION	Sequence 14 from patent US 5670367.	7218 bp DNA		
ACCESSION	166494			
VERSION	166494.1	GI:2724471		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 7218)			
TITLE	Dorner, F., Scheifflinger, F. and Falkner, F. Gunter.			
JOURNAL	Recombinant fowlpox virus Patent: US 5670367-A 14 23-SEP-1997;			
FEATURES	Location/Qualifiers 1..7218			
source	/organism='unknown'			
BASE COUNT	1944 a 1491 c 1486 g 1929 t			368 others
ORIGIN				

[illegible]

OY 401 TPAACATGAAAAAGTTATGTCGAGAGATATCTATTAAGATCAGATCAAGGTG 460
DB 1174 RRR 1115
OY 461 GACAGCAGTGTAGTTAAGCAGATTAACATTAATGTCAGAGTACATAGCAA 515
DB 1114 RRR 1060
RESULT 9
AC007926 155204 bp DNA HTG 17-JUL-2001
LOCUS Trypanosoma brucei chromosome II clone RCI93-3H15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 1 ordered pieces.
AC007926
AC007926 GI:14787210
VERSION HTG: HTGS_PHASE2.
KEYWORDS Trypanosoma brucei.
SOURCE Trypanosoma brucei.
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 155204)
AUTHORS El-Sayed, N.M., Ghedini, E., Song, J., Larkin, C., Wanless, D., Jones, K.,
Peterson, J., Hou, L., Zhao, H., Mason, T., Miltischer, J., Pal, G., Van
Aken, S., Uterback, T., Khalak, H.G., Gerard, C., Leech, V., Ullu, E.,
Meylille, S., White, O., Adams, M.D., Donelson, J.E., and Fraser, C.M.
Trypanosoma brucei GUTat10.1 RCI93-3H15 BAC genomic sequence
Unpublished
2 (bases 1 to 155204)
El-Sayed, N.M., Khalak, H. and Adams, M.D.
Direct Submission
Submitted (28-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Jul 17, 2001 this sequence version replaced gi:12746529.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 155204: contig of 155204 bp in length.
1 155204
Location/Qualifiers
/organism="Trypanosoma brucei"
/isolate="GUTat10.1"
/db_xref="taxon:5691"
/chromosome="II"
/clone="RCI93-3H15"
BASE COUNT 48747 a 31603 c 32329 g 42525 t
ORIGIN
Query Match 7.4%; Score 63.2; DB 2; Length 155204;
Best Local Similarity 46.2%; Pred. No. 0.041;
Matches 209; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
OY 397 AATATTAAACATGAAAAAGTTATGTCGAGAGATATCTATTAAGATCAGATCAAA 456
DB 53570 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53629
OY 457 GGTGACAGCAGTGTAGTTAAGCAGATTAACATTAATGTCAGAGTACATAGCAAT 516
DB 53630 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53689
OY 517 TATTATAGTCGACAAAGTGCATTTGAGAAAGCCCTTCCAGGTTCTAAATA 576
DB 53690 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53749
OY 577 ACTGTGATTAATGAGAAACACAAATGATGTAACATTCACAAAGGCTATGGGTATAT 636

DB 53750 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53809
OY 637 AATAGTTTTCATTAACACAAAACCAAAATTAAGATGACGACAAAAGAGTTGTT 696
DB 53810 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53869
OY 697 AATTAATCAGAGTTGTCATCAAGACAGTGTAGGAGAGAGAGAGAGAGAGAGAG 756
DB 53870 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53929
OY 757 AATCATCTGTCGACATTAATTAATGTCATCCGTTATGAGGTACTGTAAAGTCAA 816
DB 53930 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53989
OY 817 TTTAAAGTTTAAACAGATTAAGATACCAA 848
DB 53990 TACTAATATTAATTAATTAAGACGATTAATTAATTA 54021
RESULT 10
AC083827
LOCUS AC083827 208546 bp DNA HTG 02-OCT-2000
DEFINITION Homo sapiens clone RP11-345C21, LOW-PASS SEQUENCE SAMPLING.
AC083827
VERSION AC083827.1 GI:10445290
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 208546)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 208546)
Waterston, R.H.
Direct Submission
Submitted (02-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
Center project name: H_NH0345C21.
* NOTE: This record contains 253 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 1642: contig of 1642 bp in length
1 1652: gap of unknown length
1 1653 1653: contig of 478 bp in length
1 2131 2140: gap of unknown length
1 2131 2141: contig of 1231 bp in length
1 3371 3381: gap of unknown length
1 3382 3382: contig of 792 bp in length
1 4173 4183: gap of unknown length
1 4174 4184: contig of 1474 bp in length
1 5657 5667: gap of unknown length
1 5658 5668: gap of unknown length
1 5668 5670: contig of 803 bp in length
1 6471 6480: gap of unknown length
1 6481 6499: contig of 1019 bp in length
1 7500 7509: gap of unknown length
1 7510 8800: contig of 1291 bp in length
1 8801 8810: gap of unknown length
1 8811 9274: contig of 464 bp in length
1 9275 9285: gap of unknown length
1 9285 10198: contig of 914 bp in length
1 10199 10208: gap of unknown length

10209	11002: contig of 794 bp in length	47764	47773: gap of unknown length
11003	11012: gap of unknown length	47774	49126: contig of 1353 bp in length
11013	12073: contig of 1061 bp in length	49127	49136: gap of unknown length
12074	12083: gap of unknown length	49137	49976: contig of 840 bp in length
12084	12197: contig of 114 bp in length	49977	49986: gap of unknown length
12198	12307: gap of unknown length	49987	51259: contig of 1273 bp in length
12208	13363: contig of 1156 bp in length	51260	51269: gap of unknown length
13364	13373: gap of unknown length	51270	51810: contig of 541 bp in length
13374	14175: contig of 802 bp in length	51811	51820: gap of unknown length
14176	14185: gap of unknown length	51821	53886: contig of 2066 bp in length
14186	15448: contig of 1263 bp in length	53887	53896: gap of unknown length
15449	15458: gap of unknown length	53897	54709: contig of 813 bp in length
15459	16424: contig of 966 bp in length	54710	54719: gap of unknown length
16425	16434: gap of unknown length	54720	55769: contig of 1050 bp in length
16435	17475: contig of 1041 bp in length	55779	55779: gap of unknown length
17476	17485: gap of unknown length	55780	56320: contig of 541 bp in length
17486	18253: contig of 768 bp in length	56321	56330: gap of unknown length
18254	18263: gap of unknown length	56331	57713: contig of 1383 bp in length
18264	19283: contig of 1020 bp in length	57714	57723: gap of unknown length
19284	19293: gap of unknown length	57724	58504: contig of 781 bp in length
19294	20134: contig of 841 bp in length	58505	58514: gap of unknown length
20135	20144: gap of unknown length	58515	59576: contig of 1062 bp in length
20145	21630: contig of 1486 bp in length	59577	59586: gap of unknown length
21631	21640: gap of unknown length	59587	60391: contig of 805 bp in length
21641	22167: contig of 527 bp in length	60392	60401: gap of unknown length
22168	22177: gap of unknown length	60402	61616: contig of 1215 bp in length
22178	23135: contig of 958 bp in length	61617	61626: gap of unknown length
23136	23145: gap of unknown length	61627	63058: contig of 1432 bp in length
23146	24505: contig of 1360 bp in length	63059	63068: gap of unknown length
24506	24515: gap of unknown length	63069	64335: contig of 1287 bp in length
24516	26306: contig of 1791 bp in length	64356	64365: gap of unknown length
26307	26316: gap of unknown length	64366	64709: contig of 334 bp in length
26317	27128: contig of 812 bp in length	64710	65543: contig of 834 bp in length
27129	27138: gap of unknown length	65544	65553: gap of unknown length
27139	28060: contig of 922 bp in length	65554	66485: contig of 932 bp in length
28061	28070: gap of unknown length	65554	66485: gap of unknown length
28071	29460: contig of 1390 bp in length	66486	66495: gap of unknown length
29461	29470: gap of unknown length	66496	66743: contig of 248 bp in length
29471	30856: contig of 1386 bp in length	66744	66753: gap of unknown length
30857	30866: gap of unknown length	66754	67732: contig of 979 bp in length
30867	31404: contig of 538 bp in length	67733	67742: gap of unknown length
31405	31414: gap of unknown length	67743	68475: contig of 733 bp in length
31415	32425: contig of 1011 bp in length	68476	68485: gap of unknown length
32426	32435: gap of unknown length	68486	69665: contig of 1180 bp in length
32436	33719: contig of 1284 bp in length	69666	69675: gap of unknown length
33720	33729: gap of unknown length	69666	70517: contig of 842 bp in length
33730	34823: contig of 1094 bp in length	70518	70527: gap of unknown length
34824	34833: gap of unknown length	70528	71693: contig of 1166 bp in length
34834	35826: contig of 993 bp in length	71694	71703: gap of unknown length
35827	35836: gap of unknown length	71704	73081: contig of 1378 bp in length
35837	36653: contig of 817 bp in length	73082	73091: gap of unknown length
36654	36663: gap of unknown length	73092	73916: contig of 825 bp in length
37334	37343: gap of unknown length	73917	73926: gap of unknown length
37344	37531: contig of 188 bp in length	73927	74769: contig of 843 bp in length
37532	37541: gap of unknown length	74770	74779: gap of unknown length
37542	38960: contig of 1419 bp in length	74780	75081: contig of 302 bp in length
38961	38970: gap of unknown length	75082	75091: gap of unknown length
38971	39778: contig of 808 bp in length	75092	75731: contig of 640 bp in length
39779	39788: gap of unknown length	75732	75741: gap of unknown length
39789	41572: contig of 1784 bp in length	75742	76481: contig of 740 bp in length
41573	41582: gap of unknown length	76482	76491: gap of unknown length
41583	42900: contig of 1318 bp in length	77492	77493: contig of 1002 bp in length
42901	42910: gap of unknown length	77494	77503: gap of unknown length
42911	43747: gap of 837 bp in length	77504	78239: contig of 736 bp in length
43748	43757: gap of unknown length	78240	78249: gap of unknown length
43758	44981: contig of 1224 bp in length	78250	79249: contig of 1000 bp in length
44982	44991: gap of unknown length	79250	79259: gap of unknown length
44992	45495: contig of 504 bp in length	79260	79751: contig of 492 bp in length
45496	45505: gap of unknown length	79752	79761: gap of unknown length
45506	46565: contig of 1060 bp in length	79762	80611: contig of 850 bp in length
46566	46575: gap of unknown length	80612	80621: gap of unknown length
46576	47763: contig of 1188 bp in length	80622	81941: contig of 1320 bp in length
		81942	81951: gap of unknown length

OY	4	GATTAAATTCGAAATATCCAAATGTCACATCGATTAAGGCGATGGCCGACAGCCGGT	63
Db	63532	GATTAATTTATTTAAATTTCAATTAATTTCTTAATTAATTAAGAAACAAATGTTCTTACGGATGAT	63411
OY	64	ACAGTAAGAATAGAGGGTTATAGTAAACACAGTACCATTACTGTTTAAAGGTGACAGGTG	123
Db	63412	AATATTATTAATTAAGAAATAATTTAAATAATATGAAAAAAGAAAAAAGAGACTTATG	63471
OY	124	GGTCAAGCATTTATTACACACAGCGGTGCACAAATTACATTCATGATGATACAAAAA	183
Db	63472	TTTTTCAGATATACCTTGAAGAATATTGCGAAGAAATTAATTCAAAAGTGAAAAAAACAT	63531
OY	184	TTAACTGATGTTTCGGGATTGTCAGAAATTTGGAAGTACAAAGGAATAATTACGCAACAA	243
Db	63532	ACACCAATGTTTATGACACACATATATTAATGAAAAAATGTTAAATGAATTAATCTTTTCA	63591
OY	244	AATACTTCAGATGACCAAGTACAGTACGATTAACATCTGGGAATTAATCAACGAATGTTACG	303
Db	63592	AATGAATCAAAAATTAATGAGTGAACAAAAATATCCAAATTAAGAAAAACAAAATATATCTTTA	63651
OY	304	GTTCTAATAAGTGAAGCGGGAACAAGTAGTGTCTTCTATTAATAAACGGGAGATATGCTA	363
Db	63652	CAAAATTAATAAATGGAGATGATACAGAAAGATTAATAAATACAAAAAATAATTAATGTTT	63711
OY	364	CCAGAAAGTACGACACATGTCACGATGCTTTTAAATATTAACAATGAATAAAGTATATGTA	423
Db	63712	TCAGATGATGTCCAAATATTTTAATGAAATTCAAAATTAAGAAATTAAGAAAGGTGAAA	63771
OY	424	TCGAAGCATATTACTATTAAGATCAGATTCACGCTGACAGCAGCTTGAATTTAAGCACA	483
Db	63772	AAAAAAGGGGTACTTTTATGATGAATTTAGAAAGGACAAAAAAGTTTGATATCCAAA	63831
OY	484	TTAAACATTAAATGTACAGTACACATAGCAATTAATTAATAGTGACAAAGTGCAATTACT	543
Db	63832	AAAAATTAACATTTGAAAAAATGATTAATTAATCAATTAATACAAAAAATTAATGAATATATGAG	63891
OY	544	GATTTTGAAAAAACCTTTCCAGCTTCTAAAAATACGTGTAATATACGAAGAACACAAAT	603
Db	63892	AATACGAAAAAAGATGAACAAGATGCGTACTATTTCTTAAGAGCTCAAAAAAGAACAA	63951
OY	604	GATGTA---ACAATTCACAAAGCGATGGGTATTAATAGTTTTCATTAATACACAA	660
Db	63952	AATGATATTAATAGTATTAATAATATCTTTTCAAAATCAACCTGGAATATATTACCAAAAT	64011
OY	661	ACCAAAATTTAGCAATGAACAGCAAAAGAGATTGTTAATAATTCACAAAGCTTGATCAA	720
Db	64012	AACATAAAAAGATTAATGAATAATTAATGATTTAAAAAATTAATTAATGGAAGAAGA	64071
OY	721	GAGCATGGTAAAGGAAGAGTCAA	743
Db	64072	AAAGAAAAAGAAATCAAGAGATAAA	64094
RESULT	14		
AC004157			
LOCUS	AC004157	169546 bp	DNA
DEFINITION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING INCOMPLETE		
ACCESSION	AC004157		
KEYWORDS	HTG: HTGS_PHASE1.		
SOURCE	Plasmodium falciparum malaria parasite P. falciparum.		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	1 (bases 1 to 169546)		
AUTHORS	Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T., Kurdil, O.B., Conway, A.B. and Davis, R.W.		
JOURNAL	Plasmodium falciparum 3D7 chromosome 12		
REFERENCE	2 (bases 1 to 169546)		
AUTHORS	Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.		
JOURNAL	Direct Submission		
TITLE	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology		

COMMENT	BASE COUNT	ORIGIN
Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
On Aug 12, 2000 this sequence version replaced gi:8810447.		
* NOTE: This is a 'working draft' sequence. It currently		
* consists of 2 contigs. The true order of the pieces		
* is not known and their order in this sequence record is		
* arbitrary. Gaps between the contigs are represented as		
* runs of N, but the exact sizes of the gaps are unknown.		
* This record will be updated with the finished sequence		
* as soon as it is available and the accession number will		
* be preserved.		
* 1 23466: contig of 23466 bp in length	69871 a	15381 c 15705 g 68389 t 200 others
* 23467 23666: gap of unknown length		
* 23667 169546: contig of 145880 bp in length.		
FEATURES		
source		
1. 169546		
/organism="Plasmodium falciparum"		
/db_xref="taxon:5833"		
/chromosome="12"		
/clone="PFYAC2933"		
/clone="3D7"		

Query Match	Similarity	7.28;	Score 61.4;	DB 2,	length 169546;
Best Local	Similarity	43.68;	Pred. No. 0.083;		
Matches	324;	Conservative	0;	Mismatches 416;	Indels 3; Gaps 1
QY 4	GATTAATAATGCAAAATATACAAAATGGTGCACATGATTAATTAAGTGGCATGGCCGCAAGCGGT 63				
DB 74936	GATTAATAATATTTATTTGAATATTTCTGAAATATATGAAAAACAATGTCTTACGATGAT 74995				
QY 64	ACAGTAAAGATAGAGGGTTATAGTAAACAGTACCATTAACTGTTAAAGGTGACAGGTG 123				
DB 74996	AATATTTAATAAGAAAGAAATTTAAAAAATATCAAAAAAAGAGAGGTATAG 75055				
QY 124	GGTCAAGCAGTTATTAACACCCAGCGTCAACAATTCAATGAATGATTAAGTAGAAAAA 183				
DB 75056	TTTTGCAGATATCTTGGAAAGAAATTTGCCGAAGAAATATTCAAATGTGAAAAAACAAT 75115				
QY 184	TTAAGTATGATTTTCGGGATTTTCAGAAATTTGGAAGTTCAGAGAAAGAAATTTAAGCAACA 243				
DB 75116	ACAAGCAATGTTATGACACATATATTAATGAAAAATGTTTAATGAATTAACCTTTTCA 75175				
QY 244	AATACTTCAGATGACAAAGTACTACATGATACATCTGGGAATTAATCAACGAATGTTAG 303				
DB 75176	AATGATATCAAAATAATATGATGAGAAAAATATCCATTATGAAAAACAATAATATCTTTA 75235				
QY 304	GTTTCATTAAGTGAAGCGGGAACACAGTACTGTTTCTATTAATTAACGAGCATATGCTA 363				
DB 75236	CAAAATTAATAATGGATGATTAACAGAAAGATTAATAACAAAAAATTAATGTTT 75295				
QY 364	CCAGAAGATACACACATGACATGCTTTTAAATATTAACATGCAAAAAAGTTATGTA 423				
DB 75296	TCAGATGATGTCAAATATTTAATGAGAAATTCAAATTAAGAAATTAAGAAAAAGTGAAA 75355				
QY 424	TCGAAGAATATTACTATAAGATCAGATTCAGAGTGCACACAGTAAATTAAGCACA 483				
DB 75356	AAAAAAGGGCTAACCTTTATGATGATTAATGAGAAGAGACAAAAAGTTGATGATCCAAA 75415				
QY 484	TTAAACATTAAATGTGACAGGTACACATAGCAATTATTAATGTGACAAAGTCAATTACT 543				
DB 75416	AAAAATTAACATTTGAAAAAATATGATTAATATCAATTAAGCAAAAAATTAATATATATGAG 75475				
QY 544	GATTTTGAAGAAAGCCTTTCCAGTTCATAAATATACGTTGATATATCGAAGAACAAATT 603				
DB 75476	AATACGAAAAAAGATGAGAACAAAGTACGATCAATCTTGATGAGACGTCAAAAAAGAACAA 75535				
QY 604	GATGTA---ACAATTCACACAGCATATGGGTATATTAAGTTTTCATTAATTCACAAA 660				
DB 75536	AATGTAATTAATAGTATTAATAATATCTTTTCAAAATCAACCTGGAATTAATTTACTCAAA 75595				

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 21:25:46 ; Search time 1549.43 Seconds
(without alignments)
5888.075 Million cell updates/sec

Title: US-09-813-820-3
Perfect score: 849
Sequence: 1 GACGATATAAATGCAAAAT.....AACAGATTAAGATACCAAG 849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72.4	8.5	1007	13	CNS06X9S
2	67.4	7.9	1101	13	CNS00EVL
3	67	7.9	641	13	AO946120
4	65.2	7.7	1101	13	CNS0039G
5	64.8	7.6	1101	13	CNS0182P
6	64.6	7.6	1059	13	CNS0022B
7	64	7.5	879	13	CNS01JNG
8	63.2	7.4	1042	13	CNS0148K
9	61.4	7.2	1001	13	CNS0155H
10	60.6	7.1	893	13	CNS013XE
11	60.2	7.1	1223	13	B12981
12	60	7.1	1204	13	CNS016E2

C 13	59.8	7.0	1101	13	CNS00210	AL061917 Drosophila
C 14	59.2	7.0	700	13	AO940248	AO940248 Sheared D
C 15	58.2	6.9	997	13	CNS005TE	AL060767 Drosophila
C 16	58	6.8	1101	13	CNS0006J	AL062049 Drosophila
C 17	57.8	6.8	1101	13	CNS001FB	AL060732 Drosophila
C 18	57.8	6.8	1125	10	AL547503	AL547503
C 19	57.8	6.8	1277	11	BF264952	BF264952 HY_CEA001
C 20	57.6	6.8	729	13	AO945618	AO945618 Sheared D
C 21	57.6	6.8	959	13	CNS00655	AL062806 Drosophila
C 22	57.6	6.8	1135	13	CNS0036Q	AL226115 Tetradon
C 23	57.6	6.8	1147	13	B13042	B13042 T30M24-Sp6.
C 24	57.4	6.8	804	13	B12681	B12681 F27D1-Sp6.1
C 25	57.4	6.8	897	13	CNS07ABZ	AL46389 T7 end of
C 26	57.4	6.8	938	13	CNS006MT	AL065906 Drosophila
C 27	57.4	6.8	1091	13	CNS014AC	AL039902 Drosophila
C 28	57.2	6.7	834	13	B12387	B12387 F21E20-Sp6.
C 29	57.2	6.7	964	13	CNS07EBR	AL441457 T7 end of
C 30	57.2	6.7	1101	13	CNS014XJ	AL104737 Drosophila
C 31	57	6.7	997	13	CNS0134P	AL102403 Drosophila
C 32	56.6	6.7	945	13	CNS04DOK	AL285149 Tetradon
C 33	56.4	6.6	942	13	CNS018GS	AL109318 Drosophila
C 34	56.4	6.6	974	13	CNS001TT	AL075432 Drosophila
C 35	56.4	6.6	1885	10	BE420745	BE420745 HMM002.B0
C 36	56.2	6.6	915	11	BG520365	BG520365 ps20C09.Y
C 37	56	6.6	907	13	CNS021J4	AL176953 Tetradon
C 38	55.8	6.6	895	13	CNS06ESV	AL396821 T7 end of
C 39	55.4	6.5	868	10	AL514901	AL514901
C 40	55.4	6.5	846	13	CNS00G25	AL072931 Drosophila
C 41	55.4	6.5	994	13	CNS04NOJ	AL298972 Tetradon
C 42	55.4	6.5	1101	13	CNS00088	AL063632 Drosophila
C 43	55.2	6.5	1034	13	CNS010B4	AL098746 Drosophila
C 44	55.2	6.5	1101	13	CNS008K2	AL057309 Drosophila
C 45	55	6.5	806	13	CNS04ABE	AL281759 Tetradon

ALIGNMENTS

RESULT 1	CNS06X9S	1007 bp	DNA	GSS	06-JUL-2001			
LOCUS	T3 end of clone AX0A039F08 of library AX0A from strain CBS 7064							
DEFINITION	of Pichia farinosa, genomic survey sequence.							
ACCESSION	AL419462	GI:12202640						
VERSION	AL419462.1							
KEYWORDS	GSS.							
SOURCE	Pichia farinosa.							
ORGANISM	Pichia farinosa							
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.							
AUTHORS	1 (bases 1 to 1007) de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B., Wincker,P., Artiguenave,F. and Potier,S.							
TITLE	Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila							
JOURNAL	FEBS Lett. 487 (1), 87-90 (2000)							
MEDLINE	20584725							
REFERENCE	2 (bases 1 to 1007)							
AUTHORS	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durans,P., Lepingle,A., Llorente,B., Malpertuy,A., Neugeilse,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.							
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies							
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)							
MEDLINE	20584711							
REFERENCE	3 (bases 1 to 1007)							
AUTHORS	Genoscope.							
TITLE	Direct Submission							
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :							

COMMENT
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactic* var. *lactic*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
source
Location/Qualifiers
1..1007
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0A039F08"
/clone_1ib="AX0A"
/note="end : 13"

BASE COUNT
533 a 86 c 126 g 161 t 101 others

ORIGIN

Query Match
Best Local Similarity 8.5%; Score 72.4; DB 13; Length 1007;
Matches 217; Conservative 65; Mismatches 306; Indels 0; Gaps 0;

OY 145 GACGGTGAACAATTCATTCATTAAGTAAAAATTAGTGTGCGGATT 204
DB 54 GATGCGTGTGTATATAATTAAGMAAAAGMAAAAGMAAAAGMAAA 113
OY 205 GCAGATTGAGTACAGAGAAATTTACCCAAACAAATCTTCAGATGACAAAGTA 264
DB 114 AAAMWMAAGGCATTCGCTCTCTGAGGAAAAAAATWMAAAWMTAAAMMAAAAT 173
OY 265 GGTACATATACATCTGGAAATAATCAAGAAATGTTACGTTCAATAAAGTGAAGCGGA 324
DB 174 ATAAAGTTTAAAAAGAAAGAAATGAAWAAWMTAAAMWMAAAACMGTRCATGTW 233
OY 325 ACAAGTAGTGTCTTATTAATAACGGAGATATGCTACAGAAATACGACATGTA 384
DB 234 CMWGGTATGATGAMGTATWMTWMAAAABAAWRTAAWGTAAAGMAAAWMAAAW 293
OY 385 CGATGCTTTTAAATTTAAATGAAGAAAGTTATGTAAGATATTACTATTAAG 444
DB 294 AGAAGWGTGTATWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAA 353
OY 445 GATCAGATTCAAGGTGACAGCAATTAGATTAAAGCATTAAACATTATGTGACAGGT 504
DB 354 AAAAAMWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAW 413
OY 505 ACAATAGCAATTTATTAAGTGAAGCAATTAAGTATTTTGAAGAAAGCTTTCCA 564
DB 414 WAAAAAAMWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAW 473
OY 565 GGTCTTAAATTAATGTTGATATAGGAAGAACAAATTTATGTAACATTTCCAGAGGC 624
DB 474 AAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAW 533
OY 625 TATGGGTATATATATATTTTCAATTAATTAACAAACCAAAATTCAGATGAAGCA 684
DB 534 AAAGATTAATAAATAATWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAA 593
OY 685 AAAGATTGTTTATTAATTCACAGCTTGTATCAAGAGCATGTGTAAG 732
DB 594 AAGCTTTAGTGGGGAATAAAAACTAAATTAATAATATCAAG 641

RESULT 2
CNS00EVL 1101 bp DNA GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR29B23 of RPL1-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
The BDCP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDCP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPL1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDCP from the isogenic strain Y2; cn bw sp, the same strain used for the BDCP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1ib="RPL1-98"
/clone="BACR29B23"
/note="end : 17"

BASE COUNT
419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match
Best Local Similarity 7.9%; Score 67.4; DB 13; Length 1101;
Matches 192; Conservative 110; Mismatches 288; Indels 7; Gaps 1;

OY 251 CAGATGACAAAGTACATACATCTGGGAATTAATCAAGATGTTACGTTTACATA 310
DB 442 CMMMMHMAATGTCAGATWMMMAATWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAW 501
OY 311 AAAGTGAAGCGGACACAGTGTGTTCTTATTAATAACGAGATATGTCACGAAG 370
DB 502 AWWMAATTTTWWMTWATTTTWWMTWATTTTWWMTWATTTTWWMTWATTTTWWMTWAT 561
OY 371 ATAGCACATGTGACG-----TGGTTTAAATTAATTAACATGAAGAAAGTTATGTA 423
DB 562 AWTATTAATAATTTAAWMAWMTATTTAATTTAATAATTTAATAATTTAATAATTTT 621
OY 424 TCGAAGATATTAATTAAGATCAAGTCAAGTGAAGCAAGTGAAGTGAAGCA 483
DB 622 TTTAATAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 681
OY 484 TTAACATTAATGTCAGGTACACATAGCATTAATTAATGTCAGCAAGTGAAGTACT 543
DB 682 TTTAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 741
OY 544 GATTTGAAGAAAGCTTTCAGTTCATAAATTAAGCTGATATATGAGAAAGCAAT 603
DB 742 AWAATWMAWMTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 801
OY 604 GATGTAACATTCACAGGCTATGGTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 663
DB 802 WAAATTAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAATTAAT 861

OY	664	AAATTTTCGATGAACGCAAGCAAGAGCTTGTATTAATTCAACAAGCTTGATCAACAG	723
Db	862	ATWATAATTAATAAAAAAATWATAWTTWTWTTTTTWAMWATITAAAWMATAWMAAAA	921
OY	724	CATCGTAGCAAGAAGCAAGCAAGCAATCATTTATCATCTGTGCACAAATTAATGCT	783
Db	922	AAAAAAAAATTAAMMMMTWATATTTWTTATTAATAATWTTATWATTWATWMAWTTATATMT	981
OY	784	AATCCGGTATTTGAAGTACTGTAAAAAGTCATTAATAAGTTTAAACAGATPAA	840
Db	982	TWATWTTATATATATWTTWTTWTTATATTTATTAAMWTTATTTTAAAMAATPATATA	1038
RESULT	3		
LOCUS	AQ946120/c		
DEFINITION	Sheared DNA-46J23, TR Sheared DNA Trypanosoma brucei genomic clone		
ACCESSION	AQ946120		
VERSION	AQ946120.1		
KEYWORDS	GSS.		
SOURCE	Trypanosoma brucei.		
ORGANISM	Trypanosoma brucei Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
REFERENCE	Trypanosoma. 1 (bases 1 to 641) El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Yilm,E., Melville,S., Donelson,J., Fraser,C. and Adams,M. Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library Unpublished (1999)		
JOURNAL	Other-GSSS: Sheared DNA-46J23.TR Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org		
COMMENT	Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tcb/mdb/lbdb/ . Seq primer: M13-Reverse Class: shotgun.		
TITLE	Location/Qualifiers		
FEATURES	Source		
1..641	/organism="Trypanosoma brucei" /strain="TREU927/4 GUTat 10.1" /db_xref="taxon:5691" /clone="Sheared DNA-46J23" /clone_1bp="Sheared DNA" /note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + l method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."		
BASE COUNT	213 a 35 c 10 g 383 t		
ORIGIN			
Query Match	7.9%; Score 67; DB 13; Length 641;		
Best Local Similarity	44.7%; Pred. No. 0.0022;		
Matches	259; Conservative 0; Mismatches 320; Indels 0; Gaps 0;		
266	CTACGATACATCTCGGGAATTAATCAACGAATTTACGTTTCATTAAGAATGACGGGAA	325	

[illegible]

RESULT	4			
LOCUS	CNS00396/c			
DEFINITION	CNS00396	1101 bp	DNA	GSS
ACCESSION	BACR08K10	of RPECI-98	library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
VERSION	AL063921			
KEYWORDS	AL063921.1	GI:4941778		
ORGANISM	GSS.			
SOURCE	Fruit fly.			
REFERENCE	Drosophila melanogaster			
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
TITLE	1 (bases 1 to 1101)			
JOURNAL	Genoscope.			
COMMENT	Direct Submission			
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPECI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp; the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or			

AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr
 - Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billard at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.

FEATURES

Location/Qualifiers
 1..1059
 /organism="Drosophila melanogaster"
 /plasmid="pBelobAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN01115"
 /note="end : 17"

BASE COUNT 14 a 21 c 66 g 760 t 198 others
 ORIGIN

Query Match 7.6% Score 64.6; DB 13; Length 1059;
 Best Local Similarity 37.0%; Pred. No. 0.0058;
 Matches 312; Conservative 64; Mismatches 466; Indels 2; Gaps 1;

5 AATAAATGGAATATACAAATGTCATGATTAAGTCAGCGCAGCAAGGGTA 64
 843 AAAAAAAAAAAAAAAAAAADAADADKATADAAAAAAAAAKBCACAAAAAACHCKCA 784
 65 CAGTAAGATAGAGGTTATAGTAAACAGTACATTAAGCTTTAAAGTGAACAGGTGG 124
 783 ACKBCKACCKKCKCWAATATAAAAMCKAKSKMBRACBCDMAAYAAACKCKMA 724
 125 GTCAAGCAGTATTACACGACGCGTGCACAAATTACATTCATGATTAAGTGAATAAT 184
 723 BACKCAAAAKMNAAAAAAAAAAAAAAAAAAKAKNKMKDKAKMNAAAAKKMAAA 664
 185 TAAAGATGTTGGGATTTGCAGATTGAAGTACAGAGAAATTTAAGCACAACA 244
 663 ADGAKAKAAATDKKKKKKAAKATGRDRAKADDAKAKAAAAAADRGAA--AA 606
 245 AATACCTCAGATGACAAAGTACGATCAATCTGGGAATTAATCAAGCAATGTTGCG 304
 605 AKAAGAGDAAGAGGCTGATRGDPAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 546
 305 TTCTATTAAGTGAACGCGGACAGTAGTCTTCTATTTATAAAGCGAGATATGCTAC 364
 545 AA 486
 365 CAGAGATGACACATGATGATGTTTAAATTAATTAACATGAAGAGTATGTAT 424
 485 AA 426
 425 CGAAGATTTACTATAAGGATCAGATTCAAGGTGACAGAGATTAGTTAAGCAGAT 484
 425 AA 366
 485 TAAACATTAATGTGACAGGTACACATAGCAATTATATGTGACAAATGCAATTACG 544
 365 AA 306
 545 ATTGTAAGAACCTTTCCAGGTTCTAATTAATCTGTATATACAGAGACACATTTG 604
 305 AA 246
 605 AATTAACAATTCACAGAGCTATGGCTATATATAGTTTCAATTAACCAAAACA 664
 245 AA 186

665 AATTAAGATGACAGCAAAAGACTTTGTTAATTAATTCACAGACTGTCATCAGAC 724
 185 AANNANNNNNNNNNAAAA 126
 725 ATGTAAGAGAGAGTGAACGGGAAATCATTTAATCATCTGTGCACAAATATTATGTA 784
 125 AA 66
 785 ATGCCGATTTGAGGTACTGTAAAGTGAATTAAGTTTAAACGACGATTAAGATA 844
 65 AANNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 6
 845 CCAA 848
 5 CTAA 2

RESULT 7
 LOCUS CNS01JRG/c
 DEFINITION Anopheles gambiae GSS T7 end of clone 14D07 of NotreDamel library
 from strain PEST of Anopheles gambiae (African malaria mosquito),
 genomic survey sequence.
 ACCESSION AL147405.1 GI:7005551
 VERSION
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 Culicoidae; Anopheles.
 1 (bases 1 to 879)
 Genoscope.
 DIRECT SUBMISSION
 Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 2 (bases 1 to 879)
 Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.
 Direct Submission
 Submitted (16-FEB-2000) BMH, Institut Pasteur, 25, rue du Dr.
 Roux, Paris 75015, France
 This clone is from an A. gambiae BAC library provided by F.H.
 Collins and sequenced by Genoscope in collaboration with the
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut
 Pasteur.

FEATURES Location/Qualifiers
 1..879
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="14D07"
 /clone_lib="NotreDamel"
 /note="end : 17"

BASE COUNT 66 a 38 c 1 g 670 t 104 others
 ORIGIN

Query Match 7.5% Score 64; DB 13; Length 879;
 Best Local Similarity 38.9%; Pred. No. 0.0077;
 Matches 271; Conservative 50; Mismatches 375; Indels 1; Gaps 1;

153 AACATTAATTCATGATTAAGTGAAGAAATTAAGTATGTTCCGGATTCAGAAAT 212
 843 AAAAAAAAAAAAAAAAAAAAAAAAAAGTAGTAATATATTTGAAATGWTATTA 784
 213 TGAAGTACAGAGAAATTTAAGCAACAATTAAGTACAGTACAGTACAGTACAT 272
 783 AAAAAAGAAAAAAAGTAAATTAATAAATAAATAAATAAATAAATAAATAAATAA 724
 273 AACATCTGGAAATTAATCAACGAATGTTACGTTCAATAAAGTGAAGCGGACAAATG 332
 723 AATATGAGMAMDKAKMNAAGATRAKRGMAAAAAAAAAAAAAAAAAATATAAAGKAA 664

OY	333	T-GTTTCTATTATTAACCGGAGGTATGTCACCAGAAGATGCACATGNCATGGT	391
Dd	663	TAAAAAAMAAAAMAAAAMAAAAMAAAARTTTTATKMKRWMGAAMAAAGAAAATTT	604
OY	392	TTTTAAATATTTACAATGAAAAAAAGTTATGATCGAAGATATTACTATAAGATCACA	451
Dd	603	GWWGTDTDAAMAAAAYTAGAAAAAAGAGAGAGKGRBARAAGAAGAKAKAAKAAAAA	544
OY	452	TTCAAGGTGACGACGACGTGATTTTAACCACTTTAACTTAATGTGACAGGTACACATA	511
Dd	543	AAAAATGTTARAAATGAAAAAAMAAAKKATRBRRAAAAAAMWAGDAAAKRMGACAAAAAR	484
OY	512	GCAATTTATTTATAGTGACAAAGTCCAAATTACTGATTTTGAAMAAAGCCTTCCAGGTTCA	571
Dd	483	AGATWTKARAAAAAAMAAAAMAAAATTDMAAAATATWMTWBAAAAAAGAKKAAGKTATTGGDR	424
OY	572	AAATPACTGTTGATPATATGAGAAACACAAATTGATGTAACAAATTCACAGGCTATGGCT	631
Dd	423	AAAAAAAAAAAAADAAAAANANNAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA	364
OY	632	CATATTAATATGTTTTTCATTTAACTAACCAAAATTCGAATGACATGACGCAAAAAGCT	691
Dd	363	AA	304
OY	692	TTGTTTAATTAATTCACAACTTGTTATCAAGAACATGTATAGGACAGACGAGGGAAT	751
Dd	303	AA	244
OY	752	CATTTAATCATACTGTGTCACATAATTAATGCTAATGCCGATATTTGAAGTACTGTAAAG	811
Dd	243	AA	184
OY	812	GTGATTTAAAGTTTTAAACAGGATTAAGATACCAA	848
Dd	183	AAATTA	147
RESULT	8		
CNS0148K/c			
LOCUS	CNS0148K	1042 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence sp6 end of BAC		
	BACN1E16 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	AL103838		
VERSION	ALI03838.1	GI:5615449	
KEYWORDS	GSS.		
SOURCE	fruit fly,		
ORGANISM	Plasmod Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1042)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefgenoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (BDGP) -		
	http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC		
	library (du Polymorphisme Humain) was made by Alain Billaud at CBPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pbelOBAC11.		

```
FEATURES
source      Location/Qualifiers
            1..1042
            /organism="Drosophila melanogaster"
            /plasmid="pBELoBAC11"
            /db_xref="taxon:7227"
            /clone_1lb="DrosBAC"
            /clone="BAC11E16"
```

BASE COUNT	68 a	55 c	41 g	690 t	188 others
ORIGIN					
Query Match	7.4%; Score 63.2; DB 13; Length 1042;				
Best Local Similarity	35.0%; Pred. No. 0.011;				
Matches 295; Conservative	91; Mismatches 455;	Indels 1;	Gaps 1;		

Query Match	7.4%	Score 63.2	DB 13	Length 1042
Best Local Similarity	35.08%	Pred. No. 0.011		
Matches 295	Conservative 91	Mismatches 455	Indels 1	Gaps
QY	8	AAAAATGGAAAAATATCAAAATGGTGCATGATTTAAAGTGGCATGGCCGACACAGCGGTACAG	67	
Db	1012	AMNAAVAVAAASRAMSAMAAAGSKAMWTAHAHSTAGCMWNSCTGCMWMCACAAAGAAVRR	9533	
QY	68	TAAAGATRAGAGGCTTAATAGTTAAACACTGACCATTAACCTGTTAAAGGTGAACAGTGGCTC	127	
Db	952	RRARRRRRRRARABGARBARRRRTAAATARRNTAAAGAACANAMWNTDSAAAGAAARSYH	8933	
QY	128	AAGCATTTATACACCGACAGCGCTCAACAATTAACATTCATGATGAAAGTGAAGAAATTA	187	
Db	892	TMAAAAAAMATMKAAAAMKACATTAATTSASAMWNTSAAIGCATTAATVTAAVSCMCMH	8333	
QY	188	GTGATGTTTCGGGATTTGCAGAAATTTGAGTACAGAGAGAAAATTTAAGCCAAACAATA	247	
Db	832	YAAATSMAGAGAAAGSMGAAAMWTVHCCAAATTARGMAAAAMGAAGSACATAABTMABTT	7733	
QY	248	CTTGAGATGACAAAGTGCTRCGATTAACATCTGGGAATTAATCAAGCATGTTACGGTTC	307	
Db	772	MMAAATASCAAAATMAMSAKCAAAAAAASSTSRCSMMMAAATAGSMAAAATGSTATTT	7133	
QY	308	ATTAAGAGAAAGGGGAAACAATGATGTTTTCTATTAAAGGGGAGATATGCTACCAAG	367	
Db	712	ACACGHAAGSATWGMAMAMCAGKATATTCATCTGCACSCATGACATRCATGTGTTTMAA	6533	
QY	368	AAGA-TACGACACATGTACGATGTTTTTAAATTTATACATGAAAAAAAGTTATGTATCG	426	
Db	652	ACGAGTGGGGAATTTTACATGTGARTHTWTAAGACATTTTGAAAAAAGAAAAAAG	5933	
QY	427	AAAGATTTACTATTAAGGATCAGATTCAGAGTGGACAGACAGTTAGTTTAACACATTA	486	
Db	592	AAAAAAAAAAAAAAAAGAAA	5333	
QY	487	AACATTTATGTGACAGGTACATGACATTAATTTATAGTGGACACATGCAATTACTGAT	546	
Db	532	AAA	4733	
QY	547	TTTGAAAAAGCCTTCCAGGTTCTTAAATACTGTTGATATACGAGAACACATTTGAT	606	
Db	472	AAA	4133	
QY	607	GTAACAATTCACAGAGCTATGGTCAATATATATAGTTTTCATTTAACTCAAAACCAA	666	
Db	412	AAA	3533	
QY	667	ATTGCGATGACAGCAAAAGAGATTGTTAATTAATTCACAGCTTGATTCAGAGCAT	726	
Db	352	AAA	2933	
QY	727	GGTAAAGAGAAAGTGAACGGGAAATCTTTATATCATCTGCTGCACATATTTAATGCTAAT	786	
Db	292	GAIAAAAAAAAAARAAA	2233	
QY	787	GCCGGTTTGAGGCTACTGTAAAAAGTGAAATTAAGCTTTTAAACACAGATTAAGATACC	846	
Db	232	AAA	1733	
QY	847	AA 848		
Db	172	AA 171		

RESULT	9
CNS0155H	
LOCUS	CNS0155H 1001 bp DNA GSS 26-JUL-1999

QY	579	TGTTGATAATTCGAAAGAACCAATTGATGACCAATTCACAAGGCGTATGTCATTTAA	638
Db	520	TATTTAANNAATTAATAATTTAAAMATTMTTAAATTTT-----TTTAATTTAATTAATA	5722
QY	639	TAGTTTTTCATTTACTACAAAACCAAAATTTAGCATGAAACAGCAAAAAGATTGTGTTAA	698
Db	573	AATATATATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	6322
QY	699	TAATTCACAAGCCTGGATCAAGGCACTGGTAAGGAAGCAAGTCAACCGGAATCATTTAA	758
Db	633	TTTAAATAAAAAAATTTGGAARGGKGGGGKTTTGGATGKRGRGATTAATAAATAAATAATA	6522
QY	759	TCATACTGTGCACAATTTAATGCTAATGCCGATATGAAGTACTGTAAATAAAGGTGATT	818
Db	693	AATATATAMTTTTTATTTTTRATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATA	7522
QY	819	AAAAAGTTTAA	830
Db	753	RATRTKTTKRDW	764

RESULT	11
B12981/c	
LOCUS	B12981 1223 bp DNA GSS 14-MAY-1997
DEFINITION	T24DII-Sp6 TAWO Arabidopsis thaliana genomic clone T24DII, DNA sequence.
ACCESSION	B12981
VERSION	B12981.1 GI:2094103
KEYWORDS	GSS.
SOURCE	thale cress:
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1223)
REFERENCE	Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.
AUTHORS	
TITLE	BAC End Sequences at ATGC
JOURNAL	Unpublished (1997)
COMMENT	Contact: Ecker J.

FEATURES	source
Location/Qualifiers	
1. .1223	
/organism="Arabidopsis thaliana"	
/strain="Columbia"	
/db_xref="taxon:3702"	
/clone="T24D11"	
/clone_11p="TAMU"	
/sex="hermaphrodite"	
/note="Vector: DELORACT; Site_1: HindIII; Site_2: HindIII"	
; Produced by Rod Wing"	
70 a 50 c 53 g 678 t 372 others	

Query Match	7.1%	Score 60.2	DB 13	Length 1223
Best Local Similarity	33.9%	Pred. No 0.039		
Matches 236	Conservative	0	Mismatches 460	Indels 0
				Gaps 0
153	AACATTACATTCATCATGATAAAGTGGAAATAATTAAGTATGTTTCGGAGATTGCGAATTT	212		
1147	AA	1088		

QY	213	TGAAGTCAAGGAAGAAATTATACGCAAAACAAATCTCAGATACACAGTAGTCAGAT	272
Db	1087	ANNBAANAAAN	1028
QY	273	AACATCTGGGAATTAATCAACAATGTTACGGTTCATAAAGTGAACGGGAACAAGTAG	332
Db	1027	NAANNANANANANAAAAAAAAANNNNNATATTAANNAAAAAAAAAAANANAAAA	968
QY	333	TGTTTCTATTATAAACGGGAGATATGCTACAGAGATACGACACATGTACGATGGTT	392
Db	967	AAANNNNNANNAAN	908
QY	393	TTTAATATTATACAAATGAAAAGTATGTATCGAAGATATCTATTAAGATCAGAT	452
Db	907	NNAAANANANNNNANAAAAAAAA	848
QY	453	TCAAGTGCACAGCAGTAGTTAGATTAAACACTTAACCTTAATGTGACAGTACACATAG	512
Db	847	NAAAAAAAAAAAAAAAAAANANANANANANAAAAAAAAAAAAAAAAANNNNNAAAAAN	788
QY	513	CAATTATTATAGTGGCAAGTGCATTACTGATTTGAAAAAGCCTTCCAGTTCTAA	572
Db	787	AAANNNNNANANANANNA	728
QY	573	AATTAAGTGTGATTAATACGAGACACACATGTGATTAACATTTCCACAGGCTATGGGTC	632
Db	727	AAAAAAAANNNNNNAAANAN	668
QY	633	ATATATAGTGTTCATTACTACAAACCAAAATTCGATGACAGACAGCAAAAAGATT	692
Db	667	ANNAANNNNNANAAAAAAAAANNNANANANAAAAAAAAAAAAAAAAAAAAAAAAAN	608
QY	693	TGTTAATTAATTCACAAGCTTGTTATCCAGAGATGTTAAGGAGACAAGTGAACGGGAATTC	752
Db	607	NNNBAANAAANANANANAAAAA	548
QY	753	ATTTAATCATACTGTGCACAAATATTATGCTAATGCGGATTTGAAGTACTGTAAAGG	812
Db	547	ANNAANANANAAANANANNNNN	488
QY	813	TGATTTAAAGTTTAAACACAGATTAAGATACCA	848
Db	487	NNNANNNANNNNNAAAAAAAAAAAAAAAAAAAAANNA	452

RESULT 12
CNS01662/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

CNS01662 1204 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL106628
AL106628.1 GI:5622852
GSS.
fruit fly.
Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1204)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pGelbac11.


```

VERSION      AQ940248.1  GI:6763513
KEYWORDS     GSS.
SOURCE       Trypanosoma brucei.
ORGANISM     Trypanosoma brucei
              Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE    1 (bases 1 to 700)
AUTHORS      El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
              Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
              Fraser, C., and Adams, M.
TITLE        Determination of clone end sequences from Trypanosoma brucei GUTat
              10.1 sheared DNA library
JOURNAL      Unpublished (1999)
COMMENT      Other_GSSs: Sheared DNA-42E21.TR
              Contact: Najib M. El-Sayed
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: nelsayed@tigr.org
              Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
              DNA library constructed at TIGR. Clones will be available for
              distribution through ATCC. Sheared DNA end sequences search page:
              http://www.tigr.org/cdb/mdb/cdbdb/.
              Seq primer: M13-Forward
              Class: Shotgun.
FEATURES     Location/Qualifiers
              1..700
              /organism="Trypanosoma brucei"
              /strain="TREU927/4 GUTat 10.1"
              /db_xref="taxon:3691"
              /clone_1lb="Sheared DNA-42E21"
              /note="Vector: pUC18; Site 1: SmaI; Constructed at The
              Institute for Genomic Research (TIGR), Rockville, MD.
              Genomic DNA isolated from a cloned population of
              Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
              sheared to give a tight size distribution (approx 2 kb).
              The v + i method used for the library construction is
              described in detail in Smith, H.O. and Venter, J.C.
              (Making small insert libraries for whole genome shotgun
              sequencing projects. In Genome Sequencing: A Practical
              Approach, eds. M. Vaubin and B. Bartell, Oxford University
              Press, 1999)."
BASE COUNT   374 a 43 c 41 g 242 t
ORIGIN
Query Match 7.0%; Score 59.2; DB 13; Length 700;
Best Local Similarity 46.6%; Pred. No. 0.065;
Matches 190; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
OY 397 AATATTAACAATAAAGGTATGATCGAAGATATCTATAAAGTCAGATTCAA 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 349
OY 457 GGTGACAGCAGTATGATTAAACATTAAATGATGACAGTACACATTAACAT 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 350 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 409
OY 517 TATTAATAGTGACAAAGTCAATCTGATTTGAAAAGCCCTTCAGGTTCTAAATA 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 469
OY 577 ACAGTGAATATCGAAGAACACAAATGATGACAAATTCACAAAGCTTGGGTCTAT 636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 529
OY 637 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 696
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 530 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 589

```

```

OY 697 AATTAATTCACAGCTTGATTCAGACATGCTAAGAGAACTGAACGGGAATCATTT 756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 590 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 649
OY 757 AATCAATCTGTCACAAATTTATGCTAATGCCGATTGACCTACT 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 650 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 697
RESULT 15
LOCUS      CNS005TE/c
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL060767
VERSION     AL060767.1  GI:4943573
KEYWORDS   GSS.
SOURCE     Fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 997)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES   Location/Qualifiers
              1..997
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_1lb="RPCI-98"
              /clone="BACR12K22"
              /note="end : TET3"
BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN
Query Match 6.9%; Score 58.2; DB 13; Length 997;
Best Local Similarity 18.0%; Pred. No. 0.095;
Matches 129; Conservative 173; Mismatches 415; Indels 0; Gaps 0;
OY 4 GATTAATAATGGAAATATACAAATGCTGACATGATTAAAGTCAGCGGCAAGCGGT 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 898 RRAAAARAAADARARARARARARARARARARARARARARARARARARARAR 839
OY 64 ACAGTGAATATCGAAGAACACAAATGATGACAAATTCACAAAGCTTGGGTCTAT 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 838 AGRGKRGRRRRARRRRARARARARARARARARARARARARARARARARARAR 779
OY 124 GGTCAACAGCTTATTCACACAGCGGTGCAACATTAACATTAAGATAAGTAGAAAA 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778 GRRRGGRGRRRRARRARARARARARARARARARARARARARARARARARAR 719
OY 184 TTAAGTATGTTTCGGGATTTGCAAGATTTGAAAGTCAAGAAATTTAAACGCAACA 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 718 GARARRRRRAAAGAGARRRRRRGRRGAGARRRRGRRGAGARRRRRRMR 659
QY 244 AATTACTTCAGATGACAAAGTACGATACATCTGGGATAAATCAAGATGTTACG 303
Db 658 TTRARRRRRRARRAGARRRRRRGRRRRRRGRRRRRAGRRRRRRARRAGA 599
QY 304 GTTCATAAAGTGAACGGGACAAAGTACTGTTTCTATTATAAAAGGAGATATGCTA 363
Db 598 RRGRRRRAGRRRRRRARRAGARRRRRRRRRRRRRRRRRRRRRRRRRRARRAGA 539
QY 364 CCAGAAGATACACACATGTACGATGTTTAAATATTACACATGAAAAAGTTATGTA 423
Db 538 GRAGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRAR 479
QY 424 TCGAAGATATTACTATAAGATGATCCAAGTGCAGACGACTTAGATTAAACACA 483
Db 478 RRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRNN 419
QY 484 TTAACATTAATGTGACAGTACACATACCAATTATTATAGTGACACAAGTCAATTACT 543
Db 418 NNNNNNNNCATANNNTTTTNTTNTNTNANNNTTTTNTTNTTNTTNTTNTT 359
QY 544 GATTTGAAAAAGCCTTCCAGGTCTAAATPACTGTTGATAATAGAGAACACAACTT 603
Db 358 NNTTTANTNTNTNNNTTNNATTTGACCTTCTATATTATTATTAATTAACMTNN 299
QY 604 GATGTAACAATCCACAGGCTATGGGTCATATATAGTTTCAATTACTACAAACC 663
Db 298 NNNNAANNNTANAAAAAANNANNNNNNNNNNNNNNNNNNNNNNNNNNNNN 239
QY 664 AAAATTACGAATGACAGCAAAAAGAGTTGTTAATATTCACAGCCTGGTATCAA 720
Db 238 AAAANNAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 182

```

Search completed: January 29, 2002, 22:42:29
 Job time: 4603 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 23:21:02 ; Search time 56.04 Seconds
(without alignments) 286.810 Million cell updates/sec

Title: US-09-813-820-4

Sequence: 1 MRGSHNNHNSDQKVAATTT.....GIEGVKGEIKVLRKDKDK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	911	81.2	1185	2 A42404	collagen adhesin -
2	194	17.3	37	2 A48620	adhesin - Staphylo
3	131	11.7	1092	2 S42798	fibronectin-bindin
4	119	10.6	336	2 D86710	collagen adhesin f
5	100	8.9	1536	2 A43855	high-molecular-we
6	99.5	8.9	352	2 A81888	probable surface f
7	98.5	8.8	4919	2 T31105	hypothetical prote
8	98	8.7	1117	2 S33851	fibronectin-bindin
9	96.5	8.6	848	2 C70203	DNA topoisomerase
10	96	8.6	392	2 A53580	neurexin III beta
11	96	8.6	704	1 S39004	finger protein MSN
12	95	8.5	359	2 E86706	hypothetical prote
13	95	8.5	666	2 A42296	lysozyme 2 (EC 3.2
14	95	8.5	2020	2 C48399	ABC-type transport
15	94.5	8.4	2902	2 C71953	toxin-like outer m
16	94	8.4	2380	2 E71604	hypothetical prote
17	93.5	8.3	532	2 T06029	hypothetical prote
18	93.5	8.3	1297	2 S39791	neurotoxin - Clost
19	93.5	8.3	2340	2 B71704	cell surface antiq
20	92.5	8.2	711	2 S73898	DNA topoisomerase
21	92.5	8.2	1943	2 B64596	toxin-like outer m
22	92.5	8.2	2910	2 T28156	DNA-directed RNA p
23	92	8.2	1037	2 T13350	transcription fact
24	92	8.2	1983	2 G86643	hypothetical prote
25	92	8.2	2500	2 S16619	hypothetical prote
26	91.5	8.2	261	2 S16619	opacity protein op
27	91.5	8.2	2150	2 S71629	sensory transducti
28	91	8.1	463	2 A44808	cellulase (EC 3.2.
29	91	8.1	2269	2 T28677	riophly protein -

30	90.5	8.1	784	2 JH0101	apolipoprotein B-1
31	90	8.0	669	2 S14535	asparagine-rich pr
32	90	8.0	858	2 A42239	adenylate cyclase
33	90	8.0	4688	2 F82885	hypothetical prote
34	89.5	8.0	416	1 A41267	transcription fact
35	89.5	8.0	454	2 T26296	hypothetical prote
36	89.5	8.0	456	2 S61327	IGA-specific metal
37	89.5	8.0	1584	2 T22674	hypothetical prote
38	89.5	8.0	1815	2 C81169	IGA-specific metal
39	89.5	8.0	2893	2 A64556	toxin-like outer m
40	89	7.9	570	2 T46261	hypothetical prote
41	89	7.9	599	2 S68118	laccase (EC 1.10.3
42	89	7.9	834	2 B82940	conserved hypothet
43	89	7.9	1166	2 T28680	fibrirogen-binding
44	89	7.9	1449	2 T30552	glucosyltransferas
45	89	7.9	1553	2 T18502	hypothetical prote

ALIGNMENTS

RESULT 1
A42404
collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C:Accession: A42404; S27665
R:Patil, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wlberg, K.; Lindberg, M.; Hook
J. Biol. Chem. 267, 4766-4772, 1992
A>Title: Molecular characterization and expression of a gene encoding a Staphylococcus
A:Reference number: A42404; MUID:92165839
A:Contents: FDA 574
A:Accession: A42404
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1185 <PAT>
A:Cross-references: EMBL:M81736
A>Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 81.2%; Score 911; DB 2; Length 1185;
Best Local Similarity 90.1%; Pred. No. 4.3e-58;
Matches 182; Conservative 2; Mismatches 16; Indels 2; Gaps 2;

QY 12 SDDRVATITSGNKSNTVTHKSEAGTSVFY-RTGDM-LPEDTTHVRFPLINNEKSYV 69
Db 144 SDDRVATITSGNKSNTVTHKSEAGTSVFY-RTGDM-LPEDTTHVRFPLINNEKSYV 203
QY 70 SKDTITKQIOGGQDLSTININVTGHSNMYSGQSAITPFEKAFPSKTYDNTKMTI 129
Db 204 SKDTITKQIOGGQDLSTININVTGHSNMYSGQSAITPFEKAFPSKTYDNTKMTI 263
QY 130 DVTIPQGGSYNSFSINKTKITNEQKPEFVNSQAWQEGKKEEVNKSFNHTVHNINA 189
Db 264 DVTIPQGGSYNSFSINKTKITNEQKPEFVNSQAWQEGKKEEVNKSFNHTVHNINA 323
QY 190 NAGIEGVKGEIKVLRKDKDK 211
Db 324 NAGIEGVKGEIKVLRKDKDK 345

RESULT 2
A48620
adhesin - Staphylococcus aureus (fragment)
C:Species: Staphylococcus aureus
C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: A48620
R:Patil, J.M.; Boles, J.O.; Hook, M.
Biochemistry 32, 11428-11435, 1993
A>Title: Identification and biochemical characterization of the ligand binding domain
A:Reference number: A48620; MUID:94032261
A:Contents: FDA 574
A:Accession: A48620

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-37 <PAT>
A:Note: sequence extracted from NCBI backbone (NCBIP:138726)

Query Match	17.3%;	Score 194;	DB 2;	Length 37;
Best Local Similarity	100.0%;	Pred. No. 3.8e-08;		
Matches	37;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;
QY	19	ITSGNKSNTVYVHKSSEAGTSSEYFYKKTDDMLPEDDTH	55	
Db	1	ITSGNKSNTVYVHKSSEAGTSSEYFYKKTDDMLPEDDTH	37	

```

RESULT      3
S42798
fibronectin-binding protein - Streptococcus "equisimilis"
C:Species: Streptococcus "equisimilis"
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S42798
R:Signaes, C.
submitted to the EMBL Data Library, December 1993
A:Reference number: S42798
A:Accession: S42798
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1092 <Sig>
A:Cross-references: NID:q436135; PIDN:CAA82330.1; PID:q436136
C:Keywords: fibronectin binding

```

Query Match	11.7%;	Score 131;	DB 2;	Length 1092;
Best Local Similarity	26.5%;	Pred. No.	0.097;	
Matches	58;	Conservative	31;	Mismatches 88;
				Indels 42;
				Gaps 13

OY	19	ITSGKSTNIVNVHSEACTS--VFYYKTGDMLEPETH---VFWPLNTIN-NEKSIVSMDI	73
Db	152	VISGRKAHEAVVIQKERGSTATDNIITYKSGDI--PTQWEDSVRMSFTTNARKSTNGYGF	209
OY	* 74	TIKDOIGGGQLD-----STLININVGTSHSNYSGSQAIDFE-----KAPFGSKITVD	123
Db	210	LVTDLTLDSTMTEVEKENLRKYAINFTGG----WIGDELLSKNGMLSLKEAENGYIWE	265
OY	124	NTKNTIDVTIQ-----GIGSY-----NSFSINYTKTIITNEOKEFVNANSQAY--OEHGKE	173
Db	266	FSGGVTSINIPEKIQDYDGSYOELNESEMOMHLVAKI----KKEVLENSSTIEYVRNESKV	321
OY	174	EVMGKSF-----NHT-----VHNINANAGIEGTAKGELKLAVL	205
Db	322	EVRGENMPIDPNSTSAIVQILRLGGMAKGIVRGAEVRIILK	360

RESULT 4
D86710
collagen adhesin [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86710
R:Bioloctin, A.; Whicker, P.; Mauger, S.; Jalllon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. In press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: D86710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <STO>
A:Cross-references: GB:AE005176; NID:g12723592; PIDN:AAK04782.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
C:Gene: lacC

Query Match	10.6%;	Score 119;	DB 2;	Length 336;
Best Local Similarity	25.2%;	Pred. No. 0.16;		
Matches 52;	Conservative 28;	Mismatches 68;	Indels 58;	Gaps 11;

```

QY 24 KSNVAVHSEASSTSV-----FYKKGDMPEDTFVRFVFLINNEKSYVK 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 EDPRVAVDMKDGSTTEVYNGPATGPDKREVLTKMGWVDENDSTLIHQVYNIAKENIK 13
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 DITIKDIOGGQOL--DLSTLINVGTGHSNY--YSGSAITBEKAFPGSKITVDN 12
   - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 131 AI-YTIDIGGNQNLVSGSISIVANVYSSDGENYNDVSYPQASI--DENGVNPFVNLID 18
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 TKRTIDVTPQIGISNBSISINFKRIIN---EQQKEVNNNSAMVQEHKEEYNKSKSN 18
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 ISMTI-----TIDYSPRAIDGGLSQO-----YENRG--ELTGENIE 22
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 -----HYVN-INANAGIEGTAKGE 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 KQVEVHTPRNNGNGNASKKLSTSGE 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
A43855
high-molecular-weight surface-exposed protein HMM1 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1995
C:Accession: A43855
R:Baronkamp, S.J., Leininger, E.
Infect. Immun. 60, 1102-1113, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable
detella pertussis.
A:Reference number: A43855; MUID:92192797
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1536 <BAR>
A:Cross-references: GB:008876; GB:M84616; NTD:q475770; PIDD:AAA20527.1; PTD:q475771
A:Note: Sequence inconsistent with the nucleotide translation
A:Note: Sequence extracted from the NCBI database (NCBIN:89235, NCBI:89239)

[illegible]

```

RESULT      6
AB1888
probable surface fibril protein NMA1200 [imported] - Neisseria meningitidis (strain Z
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: AB1888
R:Perkhill, J.; Achtmann, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
R:Holroyd, S.; Jørgensen, K.; Leather, S.; Moutle, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: AB1775; MUID:20222556

```

A:Accession: A81888
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-592 <PAR>
 A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CA84461.1; PID:g737988
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA1200

Query Match 8.8%; Score 99.5; DB 2; Length 592;
 Best Local Similarity 21.6%; Pred. No. 8.5;
 Matches 43; Conservative 29; Mismatches 90; Indels 37; Gaps 6;

QY 11 GSDDKATTTGSKNSTN-----VTYHSEAGTSSVFYKTDMLPEDTTHVWF 59
 Db 147 GANCKKNNIISDTKGLNFAKETAGTNGDTVHLNGISST-----LTPTLAGSSASHV--- 198
 QY 60 LNNNEKSYSKDITIKDQIOGGQDLSTLNINVTGTHSNVYSGSALTDEKAFEGSK 119
 Db 199 -DAGNOSTHYTRASIKDVNAGW-----NIKGVKTGSGSEMDVFTYDTVE 248
 QY 120 ITVDNTKNTIDVTIPQGYGYSNSYKTKITNEOQKEFYNNQAWYQEHGKEVNGKS 179
 Db 249 FLSDDTTITTVNVSKONGKRTVEYKIAKTSVKEKDGKLVTKR-----GKGE-NGSS 300
 QY 180 FNNHTVNNINAGIEGTVC 198
 Db 301 TDEGEGLVTAKEVIDAVNK 319

RESULT 7
 T31105
 hypothetical protein 2 - Haemophilus ducreyi
 C:Species: Haemophilus ducreyi
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31105
 R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
 J. Bacteriol. 180, 6013-6022, 1998
 A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
 A:Reference number: Z20984; MUID:99030326
 A:Accession: T31105
 A:Status: preliminary; translated from GB/EMBL/DDB
 A:Molecule type: DNA
 A:Residues: 1-4919 <MAR>
 A:Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
 C:Genetics:
 A:Gene: lspa2

Query Match 8.8%; Score 98.5; DB 2; Length 4919;
 Best Local Similarity 24.8%; Pred. No. 1.4e+02;
 Matches 53; Conservative 34; Mismatches 70; Indels 57; Gaps 12;

QY 15 KVATITGKNTVYHKSSEAGTSSVFYKTDMLPEDTTHVWFNNNEKSYSKDIT 74
 Db 1104 RIGTYAGR---DLTFNKSNAAGKS-----ELINRG-----INVKNLSDS-DVS 1146
 QY 75 IKDIOGGQDLST-----LNINVTGTHSNVYSGSALTDEKAFEGSKITVDN 124
 Db 1147 FENMMQ-SQKVDLTATKTFEAKSDIELTFKNGTHPVTLNRS--NNNEKTRNS-----EN 1199
 QY 125 TKN-----TIDVTIPQGYGYSNSYKTKITNEOQKEFYNNQAWYQEHGKE 173
 Db 1200 TKNKRSIGDLINELSDSAPAEATEAYVSSSSNINPVSYLAALGNMNS----- 1250
 QY 174 EVNGKSPNHTVNNINAGIEGTVC-LKYLKQ 206
 Db 1251 --NPHYLNATKHLTGNGMODLKKOENIKYLKQ 1282

RESULT 8
 S33851

fibronectin-binding protein precursor - Streptococcus dysgalactiae
 C:Species: Streptococcus dysgalactiae
 C>Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 15-Oct-1999
 C:Accession: S33851; S32631
 R:Lindgren, P.E.; McGavin, M.J.; Signaes, C.; Guss, B.; Gurusiddappa, S.; Hoeoek, M.;
 Eur. J. Biochem. 214, 819-827, 1993
 A:Title: Two different genes coding for fibronectin-binding proteins from Streptococ
 A:Reference number: S33850; MUID:93307299
 A:Accession: S33851
 A:Molecule type: DNA
 A:Residues: 1-1117 <LIN>
 A:Cross-references: EMBL:Z22151; NID:g288970; PIDN:CAA80122.1; PID:g581643
 A:Note: the authors translated the initiation codon TTG for residue 1 as Leu
 C:Genetics:
 A:Gene: fndB
 A:Start codon: TTG
 C:Keywords: fibronectin binding
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-1117/Product: fibronectin-binding protein #status predicted <MAT>

Query Match 8.7%; Score 98; DB 2; Length 1117;
 Best Local Similarity 23.3%; Pred. No. 24;
 Matches 51; Conservative 35; Mismatches 85; Indels 48; Gaps 11;

QY 27 NVTVHSEAGTSSVFYKTDMLPEDTTH--VRNFLNNEKSYSKDITK---DQIQ 80
 Db 157 NINYNAGSFRDVS-YAKSGVLDDEDFSGRIYNNYFNENALNPI-KDEVEFMTDLP 214
 QY 81 GGQQLDSTLN-----INVTGTH-SNYSQGSALTDEPKA-----FPSKITVDNPK 126
 Db 215 QPMWTMDIEAVKRDYVQVIRGVLGVFSKNGMIDIEAVKLGILEIFSGOKLT----- 269
 QY 127 NTIDVTIPQGYGYSN-----SFSINYKTKIT---NEOQKEFYNNQAWYQEH 170
 Db 270 ----IKTPK-WTNVNGLIPIYKPLNQTSISVKTAKITEKMRQITLEYVENDSPIKEG 324
 QY 171 GKEEVNKSFNHTVNNINAGIEGTVCGLKVLKQDKD 209
 Db 325 IEMLDKNQGFQAKVRIGSGSMILGVRRGDLKVLKFERD 363

RESULT 9
 C70203
 DNA topoisomerase I (topA) homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 16-Jul-1999
 C:Accession: C70203
 R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salberg, S.; Hanson, M.; Vu
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: C70203
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-848 <RLE>
 A:Cross-references: GB:AE001180; GB:AE000783; NID:g2688755; PIDN:AAC67161.1; PID:g268
 C:Superfamily: bacterial type I DNA topoisomerase

Query Match 8.6%; Score 96.5; DB 2; Length 848;
 Best Local Similarity 22.8%; Pred. No. 22;
 Matches 59; Conservative 41; Mismatches 84; Indels 75; Gaps 14;

QY 19 ITSGNKSTNVYHKSSEAGTSSVFY-----KTGDMLPEDTTHVWFNNNEKSY 68
 Db 457 IKKGDFTSIYKMKTSHEHTAPFRTYELASLVQKMEKIGISGSPYSYSL--ISFLLEREY 513
 QY 69 VSK-DITIKDQIOGGQDL-----STLNINVTGTH-----SNYS 103

```

Db      514 AKRLNTLTPTTIGAAVINLLEKTFPVLLLENTSNMEEKIDIAIGKLDKIKYLSFEYN 573
OY      104 GOSAITD-----FEKAFPGSKITVONTNTKIDVTIPQG-YGSY-----NSFS 144
      574 GKGLKLDVWOLEPKIDSSFEFTVIESQKI---ENKNSIVTINICKYGPYLIFKGNYS 630
OY      145 INKTKITNEOKE-----FVNSQAWYQEHGKEEVNGKS--FNHTYH-NI-----NANAG 192
      631 INAKTILEMLYKKDETEKINELEKPNILGVDPLGLNVIFFKNTIYGNITVQLGEPTHAP 690
OY      193 IEGTVKG---ELKVKODK 208
      691 QETTKGKPKKIKITAKK 709

```

```

RESULT 10
A53580
neurexin III beta precursor (clone PB794-7) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 17-Mar-1999
C:Accession: A53580
R:Ushakov, Y.A.; Hata, Y.; Ichchenko, K.; Moosaw, C.; Afendis, S.; Slaughter, C.A.; S
J. Biol. Chem. 269, 11987-11992, 1994
A:Title: Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences
A:Reference number: A53580; MUID:94216308
A:Accession: A53580
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-392 <USH>
A:Cross-references: GB:L27864
C:Keywords: alternative splicing

```

```

Query Match      8.6%; Score 96; DB 2; Length 392;
Best Local Similarity 19.6%; Pred. No. 9.1;
Matches 47; Conservative 35; Mismatches 74; Indels 84; Gaps 10;

```

```

OY      5 HHHHHGSDDKVATTGSKNTVTVHKS-----EAGTSVYVYKGTGMDLPEDTTHVR 57
      52 HEHHEFGSK-----HHSVPISIVRSPVSLRGHAGATYIFKSGGLI-----YT 96
OY      58 WFLN-----INNEKSYSKDITITDQI-----QG--GQQLDLSLNI 92
      97 WPAHDRPSTRSDRLAVGFSTYVKGDLIVRIDSAPGLDGLIHEGKIGVYVNIQTVDI 156
OY      93 NWTGTHSNYSQGSATIDPEKAFPGSKITVONTNTKIDVTIPQGYGSYNSFSIN--YKTK 150
      157 SKEETPTPNVDGKYHIVFRTRNGNATLOYD-----WPNVHEHYPTG 198
OY      151 ITNEOQKEFVNS-----QAWYQEHGKEEVNGKSFNHTVHNINANAGIEGTVGEL 201
      199 NNDNEHFQVWKKIPEKYNRPVEEMLQEKGRQ-----LTIFNQQAQALIGKRGKRL 250

```

```

RESULT 11
S39004
finger protein MSN2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YM9532.Q2c; protein YMR037c
C:Species: Saccharomyces cerevisiae
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: S39004; S52886; A48131
R:Estreich, F.; Carlson, M.
Mol. Cell. Biol. 13, 3872-3881, 1993
A:Title: Two homologous zinc finger genes identified by multiplicity suppression in a SNF1
A:Reference number: A48131; MUID:93309420
A:Accession: S39004
A:Molecule type: DNA
A:Residues: 1-704 <EST>
A:Cross-references: EMBL:L08838; NID:g349594; PIDD:AAA34806.1; PID:g349595
R:Odell, C.; Bowman, S.
Submitted to the EMBL Data Library, February 1995
A:Reference number: S52885

```

```

A:Accession: S52886
A:Molecule type: DNA
A:Residues: 1-704 <CODE>
A:Cross-references: EMBL:Z48502; NID:g695715; PIDD:CAA8403.1; PID:g695717; GSPDB:GNO
C:Genetics:
A:Gene: SGD:MSN2; MIPS:YMR037c; MIPS:YMR037c
A:Cross-references: SGD:S0004640; MIPS:YMR037c
A:Map position: 13R
C:Superfamily: finger protein MSN2
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:649-665/Region: zinc finger CCH motif
F:678-698/Region: zinc finger CCH motif

```

```

Query Match      8.6%; Score 96; DB 1; Length 704;
Best Local Similarity 25.4%; Pred. No. 19;
Matches 64; Conservative 31; Mismatches 97; Indels 60; Gaps 13;

```

```

OY      12 SDRKVAITTSGNKSTVYVHKSAGTSSVYRTGMDLPEDTTHVRWFLINNEKSYSK 71
      229 SDNYSYNSISNSNSNSTGMIN---SSYFNSINIDSMLDVYSDLLDDDDDDTNLSR 284
OY      72 ---DITIKDIOG-----GQQLD-----STLNINVTGTHSNYSQGSAT 109
      285 RRSQV--ITNQPSMTNSRNSISHSLDMNHPKINPSNRNTNLIT--TNSTSSNAPNT 342
OY      110 DFEKAFPGSKITVONTNKN---TIDVTIPQGYGSYN-----SFSI 145
      343 TTMNANADSNIT-AGNPKKNDATIDNELTQILNRYNNFNNDLGTSGKKKSAKPSSEFA 401
OY      146 NYTKITNEOQ-KEFVNSO--AWYQEHGKEEVNGKSFNHTVHNINANAG---IEGTVG 199
      402 NMTKINPSQLOQOOLNRYVQHKQLTSSHNNSSTNKSFSNLDYSRQRASLPITDLSLY 461
OY      200 ELKVKODKDTK 211
      462 DL-VNKKDDEPK 472

```

```

RESULT 12
E86706
hypothetical protein ygea [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: E86706
R:Boletín, A.; Winkler, P.; Manger, S.; Jallón, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: E86706
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-359 <STO>
A:Cross-references: GB:AE005176; NID:g12723558; PIDD:AAK04751.1; GSPDB:GNO0146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ygea

```

```

Query Match      8.5%; Score 95; DB 2; Length 359;
Best Local Similarity 21.7%; Pred. No. 9.6;
Matches 58; Conservative 36; Mismatches 89; Indels 84; Gaps 12;

```

```

OY      19 ITSGNS---TNTVHKSAGTSSVYRTGMDLPEDTTHVRWFLINNEKSYVSDIT 75
      101 LTYMNSPDTYSVFKKDLDDKNVVGKTFEKKVTDRFETS--INADEISF-----FTS 154
OY      76 KQDIOGQQLDLSLNI-----NVTGTH---SNYSQGSATIDFEKAFPGSKITVONTK 126
      155 KIDAEIKKGDITVIVKRPNIIIGVITMSGVQSKKNINQHSN---NNESNSNTYK 212
OY      127 -----NTIDVTIPQGYGSYNSFSINYKTKITNEOQKEFV 160

```

```

Query Match Similarity      8.5%; Score 95; DB 2; Length 666;
Best Local Similarity      24.0%; Pred. No. 21;
Matches 49; Conservative 28; Mismatches 81; Indels 46; Gaps 11;

OY 20 TSGNSTAVTVHKSAGTSSVYRYTKGDMLPEDTTH-----VRPFLNINNEKSYVS 70
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 242 TSGNNGGSATP-----TGTTVTVKSGDSV-WGISHSFGITMAOLIEW-NNKKNNETIYG 293
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

OY 71 KDITFKDDIOGGQOQLDLSTLNI-----NVYG-THSNYSGQSAIATDEKAFPGSKITV 122
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 294 OKLTK-----GGQASGSSPTTNGNNASSCGTSGNTSGSTGQAT-----GAKYIV 340
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

OY 123 DNTKNTIDVITPGCGSTNSFSINRKTKITNEQOKEFVANSQAMQEHCKEYVNGSFMH 182
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 341 KSGDSVVKMIANDHG-----ISMQLIE-WNNKKNFVYGGQQLVYKSSSSASGTSNT 393
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

OY 183 TVHNINANAGIEG-TVKGLKVLK 205
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 394 STGNTSSMTANTGSTTSGSTYVK 417
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 14
C48399
ABC-type transport protein ydbA.2 - Escherichia coli
C:Species: Escherichia coli
C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999
C:Accession: C48399; D64891; H64891
R:Moszer, I.; Glaser, P.; Panchin, A.
Biochimie 73, 1361-1374, 1991
A:Title: Multiple IS insertion sequences near the replication terminus in Escherichia
A:Reference number: A48399; MUID:92190338
A:Accession: C48399
A:Molecule type: DNA
A:Residues: 464-2020 <MOS>
A:Cross-references: GB:D85081; NID:93041754
A:Experimental source: strain K-12
A:Note: sequence inconsistent with the nucleotide translation
R:Blahtner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617

```

```

Query Match Summary          8.5%; Score 95; DB 2; Length 2902;
Best Local Similarity 25.1%; Pred. No. 84;
Matches 69; Conservative 30; Mismatches 92; Indels 84; Gaps 16;

QY 5 HHHHHG-----SDDKVAITSGNK-----STNVYHKSEA 35
      |||::: |||
      ||::: |||
Db 94 HHHRRNSLPPTPPDESDDPVPPTPGGDEITPDDPDPPTPKPVSFNNDVILDKTEK 153
      |||::: |||
QY 36 GTS---SVPFY---KTGMLPEDTTHVRWFLNI-----NNEKSY--VSKDITTKDQIG 81
      |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 154 TLITDSVFTTYENADGFTISLQDSNGRKATITNMQIDENATVALEGVASDGAITKQIYNH 213
      |||::: |||::: |||::: |||::: |||::: |||::: |||
QY 82 GOOL-----DLSTLIN-----VTGTHSNYSGGSAITDFEK-----AFPGSK 119
      |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 214 NCELVTIGDNDATVNNRNGKTVDGKSTGTETGNGNGK--VIQGDLDVSGGSGHIDITKDS 272
      |||::: |||::: |||::: |||::: |||::: |||::: |||
QY 120 ITVYDTRKTITDITIQGG---SYNSFSINY--KTKINQOKEPVNNSQAMQYEHKKEE 174
      |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 273 AVVDN-KGTMVTDDESMGIQIDGKAIVNNEGSESTITNGGTGTQIINGDDATPANNKXTT 331
      |||::: |||::: |||::: |||::: |||::: |||::: |||
QY 175 YNGKSFNHTVHNINANAGIEGTVGKELKVLKODKD 209
      |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 332 VDGKSTGT--EINGNG-----KVI-QDDG 354
      |||::: |||::: |||::: |||::: |||::: |||::: |||

RESULT 15
C71953
toxin-like outer membrane protein jhp0274 - Helicobacter pylori (strain j99)
C:Species: Helicobacter pylori
A:Variety: strain j99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: C71953
R:Alm, R.A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Dolj, P.C.; Smith, D.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voyis, G.F.
Nature 397, 176-180, 1999
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A11800; MUID:99120557
A:Accession: C71953
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2902 <ARN>
A:Cross-references: GB:AE001464; GB:AE001439; NID:94154789; PIDN:AAD05855.1; PID:9415
C:Genetics:
C:Experimental source: strain j99
A:gene: jhp0274

Query Match          8.4%; Score 94.5; DB 2; Length 2902;
Best Local Similarity 23.1%; Pred. No. 1.4e+02;
Matches 50; Conservative 26; Mismatches 63; Indels 77; Gaps 11;

QY 23 NKSTVTVYHKSAGTSVFYKTKTGMLPEDTTHVRWFLNINNEKSYSKDITIKQ----- 78
      |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1335 NLSQNASITASHHSTLEL---QGDLINIDTSS---LNLQSAIIVSNNAITINDYASLI 1386
      |||::: |||::: |||::: |||::: |||::: |||::: |||
QY 79 IGGGQOLD-----LSTLINI-----NVLTGTHSNYSGGSAITD- 110

```

```
Db 1387 ASNGSHLNGAVNENSANITTSLSSSIVFKGAVSLRGOFNLNNSSLDFOGSSAITSN 1446
QY 111 -----PEKAPGSKITVDNWKNTIDYTIPOGYGS-----YNSFSINYKTKITNEQOKEFY 160
Db 1447 TAFNFYDNAFSOSPITF---HQALDIKVPUSLGCNLLNPNSSVNLNK----- 1491
QY 161 NNSQAWYQEHGK-----EEVNGKSFNHTVHNT 187
Db 1492 -NSQVFSDDGSLNLIANIIDLSDLNGN--KNRYYNI 1524
```

Search completed: January 30, 2002, 00:03:33
Job time: 2551 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 22:44:32 ; Search time 67.56 Seconds
(without alignments)
231.342 Million cell updates/sec

Title: US-09-813-820-4

Perfect score: 1122
Sequence: 1 MRGSHHHHSGSDKVAITIT.....GIEGVKGEKLVKQDKDPTK 211

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq..1101.*
1: /SIDS8/gcgdata/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1122	100.0	211	19	AAW31553
2	1054.5	94.0	512	19	AAW31554
3	904	80.6	1185	13	AAW31552
4	830	74.0	159	19	AAW31552
5	178	15.9	458	21	AAW31552
6	165.5	14.8	146	21	AAW31552
7	125.5	11.2	345	19	AAW31555
8	116	10.3	473	19	AAW68400
9	114.5	10.2	2032	20	AAW00241
10	114.5	10.2	2032	20	AAW00238
11	114.5	10.2	2032	20	AAW00240

12	114.5	10.2	2032	20	AAW00242	Enterococcus faeca
13	113.5	10.1	184	20	AAW32917	E.coli optimised K
14	113.5	10.1	184	20	AAW10309	Human KGF-2 constr
15	113.5	10.1	184	22	AAW61673	PGE60-Cys37 constr
16	110.5	9.8	1338	14	AAW41731	High molecular wtl
17	110.5	9.8	1598	18	AAW30291	Non-typable Haemo
18	110	9.8	493	12	AAW30992	P. falciparum spor
19	110	9.8	2314	22	AAW69136	M. catarrhalis les
20	109	9.7	1536	18	AAW30283	Non-typable Haemo
21	105.5	9.4	342	14	AAW39498	Human apoAIV mutel
22	104.5	9.3	1529	14	AAW41732	High molecular wtl
23	103.5	9.2	139	19	AAW31556	Fibronectin-bindin
24	101.5	9.0	847	17	AAW06400	Class A starch bra
25	101.5	9.0	1601	18	AAW30292	Non-typable Haemo
26	100	8.9	537	7	AAW60452	Sequence of the As
27	100	8.9	1095	21	AAW01847	Haemophilus influe
28	100	8.9	1536	14	AAW41723	High molecular wtl
29	100	8.9	1536	14	AAW41725	High molecular wtl
30	100	8.9	1536	15	AAW63505	Haemophilus influe
31	100	8.9	1536	21	AAW01846	Haemophilus influe
32	99.5	8.9	389	20	AAW23745	A surface protein
33	99.5	8.9	589	22	AAW06173	N. meningitidis p2
34	99.5	8.9	592	22	AAW06180	N. meningitidis p2
35	98.5	8.8	346	14	AAW39493	Human apoAIV mutel
36	98.5	8.8	351	19	AAW56000	Fibroblast growth
37	97.5	8.7	195	19	AAW59053	His rrf10 protein
38	97	8.6	437	19	AAW37738	Nucleotide sequenc
39	97	8.6	451	19	AAW68398	Clostridium botuli
40	96.5	8.6	1155	22	AAW82343	S. epidermidis ope
41	96.5	8.6	1802	21	AAW83170	Cell wall protein
42	96.5	8.6	1802	21	AAW70119	Staph. epidermidis
43	96	8.6	329	19	AAW44778	Human recombinant
44	96	8.6	667	21	AAW29654	Human membrane-ass
45	95	8.5	666	17	AAW85291	Enterococcus hirae

ALIGNMENTS

RESULT 1	
AAW31553	AAW31553 standard; Protein; 211 AA.
XX	
AC	AAW31553:
XX	
DT	21-MAY-1998 (first entry)
XX	
DE	Collagen binding protein M31 epitope.
XX	
KW	Collagen binding protein; cna gene; sepsis; infection;
KW	microbial surface component recognising adhesive matrix molecule;
KW	MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;
KW	epitope M31.
XX	
OS	Staphylococcus aureus.
XX	
PH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..12
FT	/note="vector POE30-derived peptide"
FT	13..211
FT	/note="epitope M31"
XX	
PN	W09743314-A2.
XX	
PD	20-NOV-1997.
XX	
PF	14-MAY-1997; 97WO-US08210.
XX	
PK	16-MAY-1996; 96US-0017678.
XX	
PA	(UABR-) UAB RES FOUND.
PA	(TEXA) UNIV TEXAS A & M SYSTEM.
XX	

PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 XX WPI; 1998-008801/01.
 DR N-PSDB; AAT93437.
 XX
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 31; Page 115-116; 143pp; English.
 XX
 CC This protein comprises Staphylococcus aureus collagen binding
 CC protein (CBP) epitope M31, i.e. amino acids 61-343 of full-length
 CC CBP, plus a vector-derived N-terminal peptide. Claimed 441, 849
 CC and 1500 bp nucleic acid sequences (see AAT93436-38) respectively
 CC encode CBP epitopes M17, M31 and M55 (see AAW31552-54) that confer
 CC protection against S. aureus infection. These nucleic acid
 CC sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. The CBP
 CC epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of S.
 CC aureus infection.
 XX
 SQ Sequence 211 AA:
 Query Match 100.0%; Score 1122; DB 19; Length 211;
 Best Local Similarity 100.0%; Pred. No. 1.2e-86;
 Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRCGSHHHHNSDDKVAATITSGNKSSTNVYHKSEAGTSSVFYKTDGMLPEDTTHVRWFL 60
 DB 1 mrgshhhhhgsddkvalitsgnksstnvvhkseagtsvfyktdgmlpedtthvrwfl 60
 QY 61 NINNEKSYVSKDTITDQIOGGQDLSTLNTINVTGTHSNYVSGSATIDPEKAFPSKTI 120
 DB 61 nlnneksyvsdkdltldqig99gqldstlninvgtshnysvgsatldfekaafpsk 120
 QY 121 TVDNTKNTIDVTIPQGYGSYNSFSINVKTKITNEOQKEFVNNSOAMQOEHGKEEVNGKSF 180
 DB 121 tvdntkntldvtlpggygsynsfisnykktikneqkefvnnsqamvqehgkeevngksf 180
 QY 181 NHTVHNINANAGIEGTVGKGLKVLKODKDTK 211
 DB 181 nhtvhninanaiegtevgkglkvlkqdkdtk 211

FT FT /note="vector pQE30-derived peptide"
 FT Protein 13.512
 FT /note="epitope M17"
 XX
 XX
 PN WO9743314-A2.
 XX
 XX
 PD 20-NOV-1997.
 XX
 XX
 PF 14-MAY-1997; 97WO-US08210.
 XX
 PR 16-MAY-1996; 96US-0017678.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (TEXA) UNTV TEXAS A & M SYSTEM.
 PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 DR WPI; 1998-008801/01.
 DR N-PSDB; AAT93438.
 XX
 XX
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 31; Page 117-119; 143pp; English.
 XX
 CC This protein comprises Staphylococcus aureus collagen binding
 CC protein (CBP) epitope M55, i.e. amino acids 30-531 of full-length
 CC CBP, plus a vector-derived N-terminal peptide. Claimed 441, 849
 CC and 1500 bp nucleic acid sequences (see AAT93436-38) respectively
 CC encode CBP epitopes M17, M31 and M55 (see AAW31552-54) that confer
 CC protection against S. aureus infection. These nucleic acid
 CC sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. The CBP
 CC epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of S.
 CC aureus infection.
 XX
 SQ Sequence 512 AA:
 Query Match 94.0%; Score 1054.5; DB 19; Length 512;
 Best Local Similarity 64.7%; Pred. No. 1.8e-80;
 Matches 211; Conservative 0; Mismatches 0; Indels 115; Gaps 1;
 QY 1 MRCGSHHHHNSDDKVAATITSGNKSSTNVYHKSEAGTSSVFYKTDGMLPEDTTHVRWFL 60
 DB 1 mrgshhhhhgsardlstlnvtldtvspskiedgkltvkmfdknkqlngdmikvwaw 60
 QY 12 -----SDKVAATITSGNKSSTNVYHKSEAGTSSVFYKTDGMLPEDTTHVRWFL 65
 DB 61 ptesgtvkiegyktvpltvkqegvgavltprdgatltfndkvekladvsgfaefevgyn 120
 QY 12 -----SDKVAATITSGNKSSTNVYHKSEAGTSSVFYKTDGMLPEDTTHVRWFL 65
 DB 121 ltgtntsdskvalitsgnksstnvvhkseagtsvfyktdgmlpedtthvrwfl 180
 QY 66 KSYVSKDITRKDQIOGGQDLSTLNTINVTGTHSNYVSGSATIDPEKAFPSKTI 125
 DB 181 ksyvskdiltldqig99gqldstlninvgtshnysvgsatldfekaafpskltvdt 240
 QY 126 KNTIDVTIPQGYGSYNSFSINVKTKITNEOQKEFVNNSOAMQOEHGKEEVNGKSF 185
 DB 241 kntidvtlpggygsynsfisnykktikneqkefvnnsqamvqehgkeevngksf 300
 QY 186 NINANAGIEGTVGKGLKVLKODKDTK 211

|||||
 Db 301 nlmanagiegvtvkgelkvkqdkdk 326

RESULT 3
 AAR22675
 ID AAR22675 standard; Protein; 1185 AA.

XX AAR22675;

XX 04-DEC-1992 (first entry)

XX Collagen binding protein.

XX CBP; collagen binding protein; mastitis; arthritis.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

XX Region 30..1185 /note="claim 11; page 29-30"

XX Peptide 1..29 /label="sig_peptide

XX Region 30..534 /label="A

XX /note="see CC"

XX Region 535..721 /label="B1

XX /note="see CC"

XX Region 722..908 /label="B2

XX /note="see CC"

XX Region 909..1095 /label="B3

XX /note="see CC"

XX Region 1096..1159 /label="W

XX /note="see CC"

XX Region 1160..1179 /label="M

XX /note="see CC"

XX Region 1180..1185 /note="charged C-terminal"

XX Region

XX WO9207002-A.

XX 30-APR-1992.

XX 22-OCT-1991; 91WO-SE00207.

XX 22-OCT-1990; 90SE-0003374.

XX (ALFA) ALFA LAVAL AGRI INT AB.

XX Guss BM, Hook M, Jonsson H, Lindberg KM, Patti J, Signaes LC;

XX Switalski LM;

XX WPI; 1992-167099/20.

XX N-PSDB; AAQ24123.

XX Hybrid DNA molecule encoding S.aureus collagen binding protein -

XX protein is expressed in E.coli and used for diagnosis e.g. of

XX septic arthritis

XX Disclosure; Fig 2; 40pp; English.

CC to mediate the binding of the protein to the cell wall. The amino

CC acid sequence nearest to the C-terminal end consists of a long

CC stretch of hydrophobic residues followed by some charged amino acids

CC This region is called M.

CC The CBP can be used for immunisation pref. in combination with a

CC fusion protein, e.g. for vaccination of ruminants against mastitis

CC caused by staphylococcal infections. It can also be used to block

CC infection in an open skin wound, e.g. for blocking protein receptors

CC or by immunisation. In the latter, the host produces specific

CC antibodies which block the adherence of the bacterial strains to

CC damaged tissue. This treatment can be used for septic arthritis

CC and tissue damage of e.g. skin, connective tissue, and mucous

CC membranes. Dosage for immunisation is 0.5-5 microg CBP/kg; for

CC topical admin. the protein is used at a concn. of 25-250 microg/ml.

XX Sequence 1185 AA;

Query Match 80.6%; Score 904; DB 13; Length 1185;
 Best Local Similarity 89.6%; Pred. No. 2,4e-67;
 Matches 180; Conservative 3; Mismatches 16; Indels 2; Gaps 2;
 QY 13 DDKVVATTSGNKSNTVHRSKSEAGTSVFY-RTGDM-LPEDTHVRPFINNERSYVS 70
 Db 145 ddkvatltsgnksntvgykvrkrepvflnksqkcygedtlhvflninnexyvs 204
 QY 71 KDITIKDQIQGGQQLDSTLININVTGHSNYSQGSATDTEKAFPGSKITVDNFKNTID 130
 Db 205 kditiKdei gggqql dsltinivtgshnysqgsatdteka fpskiltvdntkntid 264
 QY 131 VTTPQGYGSYNSFSINKTKITNEQKEFYNNNSQAWQHEGKEVNNKSNHTVHNINAN 190
 Db 265 vtTpqgygsynsfisinktkitneqgketyvnsqawygehngkeevngksinhvhninan 324
 QY 191 AGIEGTWKGELKVKODKDK 211
 Db 325 agIegtWkgelkvkqdkdk 345

RESULT 4

AAW31552
 ID AAW31552 standard; Protein; 159 AA.

XX AAW31552;

XX 21-MAY-1998 (first entry)

XX Collagen binding protein M17 epitope.

XX Collagen binding protein; cna gene; sepsis; infection;

XX microbial surface component regionising adhesive matrix molecule;

XX MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;

XX epitope M17.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

XX Peptide 1..12 /note="vector pQE30-derived peptide"

XX Protein 13..159 /note="epitope M17"

XX WO9743314-A2.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US08210.

XX 16-MAY-1996; 96US-0017678.

XX (UABR-) UAB RES FOUND.

XX (TEXA) UNIV TEXAS A & M SYSTEM.

PI Hook M, House-Pompeo K, Patil JM, Sthanam N, Symersky J;
 XX WPI; 1998-008601/01.
 DR N-PSDB; AAT93436.
 XX
 XX Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 XX Claim 31; Page 114; 143pp; English.
 PS
 XX This protein comprises Staphylococcus aureus collagen binding
 CC protein (CBP) epitope M17, i.e. amino acids 151-297 of full-length
 CC CBP, plus a vector-derived N-terminal peptide. Claimed 441, 849
 CC and 1500 bp nucleic acid sequences (see AAT93436-38) respectively
 CC encode CBP epitopes M17, M31 and M55 (see AAW31552-54) that confer
 CC protection against S. aureus infection. These nucleic acid
 CC sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. The CBP
 CC epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of S.
 CC aureus infection.
 CC
 XX Sequence 159 AA:
 SQ
 Query Match 74.0%; Score 830; DB 19; Length 159;
 Best Local Similarity 96.4%; Pred. No. 2.7e-62;
 Matches 159; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
 QY 1 MRGSHHHHGGDDKAYATTSQNKSTNTVHKSEAGTSSVFYKTDMLPEDTHVRF 60
 Db 1 mrgshhhhhg-----ltsqnkstnlvnlvkhseagtsavfyktdmlpedtlhvwf 54
 QY 61 NINNEKSYSKDITTKDOIGGQGLDSTLNTNVTGTHSNYSGSALIDFEKAFPGSKI 120
 Db 55 nnnneksyskdiitkldqigggqldstlnlnvgtshnyysggsalidcfekafpgski 114
 QY 121 FVDNFKNTIDVTIPQGYGSNSFSINYSKTKITNEQKKEFVNNSQA 165
 Db 115 tvdnkntidvtlpgdygysnfsinykxkitneqkkelvnmsqa 159

RESULT 5
 AAB37667
 ID AAB37667 standard; protein; 458 AA.
 XX
 AC AAB37667;
 XX
 DT 28-FEB-2001 (first entry)
 XX
 DE Collagen binding Ace protein.
 XX
 KW Collagen binding Ace protein; antibacterial; immunostimulant; vaccine;
 KM extracellular matrix-binding protein; lamin binding; bacterial infection.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200068242-A1.
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12590.
 XX
 PR 10-MAY-1999; 99US-0133334.
 XX

PA (TEXA) UNIV TEXAS A & M SYSTEM.
 PA (TEXA) UNIV TEXAS MEDICAL SCHOOL.
 XX
 XX Rich RL, Kriekemeyer B, Owens RT, Hook M, Murray BE;
 PI Nallapareddy SR, Qin X, Weinstock GM, Singh KV, Duh R;
 XX WPI; 2000-687639/67.
 DR N-PSDB; AAC83828.
 XX
 XX New collagen-binding protein from Enterococcus, useful e.g. in
 PT protective vaccines, for diagnosis and treatment of Enterococcal
 PT infections and for screening for compounds that inhibit collagen
 PT binding by enterococci -
 XX
 PS Claim 26; Pages 119-122; 148pp; English.
 XX
 CC The present sequence is collagen binding Ace protein from Enterococcus
 CC faecalis, where Ace is adhesion of collagen from enterococcal bacteria.
 CC This protein is an extracellular matrix-binding protein, which can bind
 CC with collagens such as collagen type I and type IV and with lamins. Ace
 CC protein can be used to inhibit attachment of enterococci to collagen and
 CC to treat or prevent enterococcal infections, specifically as a vaccine.
 CC
 XX Sequence 458 AA:
 SQ
 Query Match 15.9%; Score 178; DB 21; Length 458;
 Best Local Similarity 28.3%; Pred. No. 5.9e-07;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;
 QY 18 TITSGNKSTNTVHKSEAG---TSSVFYKTDMLPEDTHVRFNINNEKSYSKDIT 74
 Db 121 tatatqrliegvtntetqterdyfkykgvdage-sngyvwflnvlhksdvredis 179
 QY 75 IDQIGGQGLDSTLNTNVTGTHSNYSGSALIDFEKAFPGSKRTVNTKNTIDVTIP 134
 Db 180 idrfgsggqqlnhesitfdvndkety--islaefeggygkldfv--tdndfnlrfy 234
 QY 135 QGYGSNSFSINYSKTKIT--NEQKKEFVNNSQAWYOEHGKEEYNGKSFNHTVINANAG 192
 Db 235 rdkarftsfivyrstltaeqghatfensydlnglnndatneknetsqv-----knvf 289
 QY 193 ISGTWVGEELV 203
 Db 290 vgeasgngnv 300

RESULT 6
 AAB37668
 ID AAB37668 standard; protein; 146 AA.
 XX
 AC AAB37668;
 XX
 DT 28-FEB-2001 (first entry)
 XX
 DE Collagen binding Ace protein collagen binding domain.
 XX
 KW Collagen binding Ace protein; antibacterial; immunostimulant; vaccine;
 KM extracellular matrix-binding protein; lamin binding; bacterial infection.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200068242-A1.
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12590.
 XX
 PR 10-MAY-1999; 99US-0133334.
 XX
 PA (TEXA) UNIV TEXAS A & M SYSTEM.
 PA (TEXA) UNIV TEXAS MEDICAL SCHOOL.
 XX

PI Rich RL, Kriekemeyer B, Owens RT, Hook M, Murray BE;
 PI Mallapareddy SR, Qin X, Weinstein GM, Singh KV, Duh R;
 XX WPI; 2000-687639/67.
 DR
 XX
 XX New collagen-binding protein from *Enterococcus*, useful e.g. in
 PT protective vaccines, for diagnosis and treatment of *Enterococcal*
 PT infections and for screening for compounds that inhibit collagen
 PT binding by *enterococci* -
 XX
 XX Disclosure; Fig 1: 148pp; English.
 PS
 XX The present invention relates to collagen binding Ace protein from
 CC *Enterococcus faecalis*, where Ace is adhesion of collagen from
 CC enterococcal bacteria (see AAC3828 and AAB3667). Ace protein is an
 CC extracellular matrix-binding protein, which can bind with collagens such
 CC as collagen type I and type IV and with lamins. Ace protein can be used
 CC to inhibit attachment of *enterococci* to collagen and to treat or prevent
 CC enterococcal infections, specifically as a vaccine. The present sequence
 CC is the collagen binding domain of Ace protein.
 CC
 XX Sequence 146 AA:
 SQ
 Query Match 14.8%; Score 165.5; DB 21; Length 146;
 Best Local Similarity 30.2%; Pred. No. 1.4e-06;
 Matches 45; Conservative 30; Mismatches 65; Indels 9; Gaps 5;
 QY . 41 FYYKTGDMLEPDTTHVRWFLINNEKSYVSKDITIKDQIGGQQLDLSTLINVTGTHSN 100
 Db . 5 ffykvgdglge-sngyrvflvnlnksdvtdedtsldrgsqgqlnkeftfdvndket 63
 QY 101 YVSGGSAITDFEKAFFGSKRTVDNTKNTIDVTIPQGYGSGNSFINKTKIT--NEQKE 158
 Db . 64 ky---lsleefegqgkldfv--lndfnlrfyrdkxarfsvfivytlsiteagqhgat 118
 QY 159 FVNSQAWYQEHGKEVNGKSFNHTVHNI 187
 Db 119 fensydingqlmqdatneksq-vknv 146
 DE Fibronectin-binding MSCRAMM derivative pCF33.
 XX
 XX Fibronectin: pCF33; collagen binding protein; sepsis; infection;
 KW microbial surface component recognising adhesive matrix molecule;
 KM MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;
 KM mastitis.
 XX
 XX Staphylococcus aureus.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..12
 FT /note= "vector pQE30-derived peptide"
 FT
 XX
 XX WO9743314-A2.
 PN
 XX
 XX 20-NOV-1997.
 PD
 XX
 XX 14-MAY-1997; 97WO-US08210.
 PF
 XX
 XX 16-MAY-1996; 96US-0017678.
 PR
 XX
 XX (UABR-) UAB RES FOUND.
 PA (TEXA) UNTV TEXAS A & M SYSTEM.
 PA
 XX

PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 XX WPI; 1998-008801/01.
 DR
 XX
 XX Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with *Staphylococcus aureus*
 XX
 XX Disclosure; Page 91; 143pp; English.
 PS
 XX
 XX This protein comprises *Staphylococcus aureus* fibronectin-binding
 CC microbial surface component recognising adhesive matrix molecule
 CC (MSCRAMM) derivative pCF33, plus a vector-derived N-terminal
 CC peptide. The invention relates to claimed nucleic acid sequences
 CC (see AAT93436-38) encoding *S. aureus* collagen binding protein (CBP)
 CC epitopes M17, M31 and M55 (see AAW31552-54) that confer protection
 CC against *S. aureus* infection. CBP protein and antigenic epitopes
 CC are contemplated for use in the treatment of pathological
 CC infections, especially to prevent bacterial adhesion to collagen.
 CC The epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of *S.*
 CC aureus infection. pCF33 and pQD (see AAW31556) were used to raise
 CC anti-MSCRAMM polyclonal antibodies used in passive immunisation
 CC against bovine mastitis.
 CC
 XX Sequence 345 AA:
 SQ
 Query Match 11.2%; Score 125.5; DB 19; Length 345;
 Best Local Similarity 30.1%; Pred. No. 0.01;
 Matches 56; Conservative 20; Mismatches 73; Indels 37; Gaps 10;
 QY 1 MRGSHHHHHNGS---DDKVAITTSGNKSTNV-----TVHKSSEAGTSVFFYKTKDM 48
 Db 1 mrgshhhhhngsmvaadapaagtdltnqlnvctvgdsqlyphqgkykloy---gfs 57
 QY 49 LPEDTTHVRWF-----LINNEKSYVS-KDITIKDQIGGQQLDLSTLINVTGTHSN 100
 Db 58 vpsnavkgydtkltvpkehlngvtstakvypimagdqvlangvidsdg--nvlyftld 114
 QY 101 YVSGGSAITDFEKA---FPGSKITVDNTKNTIDVTIPQGYGSGNSFINKTKITNEQK 157
 Db 115 yvntk-----dvkactlmp-ayldpenvkktgnvtlactgigtsta---nktvltvdyeky 166
 QY 158 EFVNS 163
 Db 167 kfylns 172
 DE Clostridium botulinum type G strain 113/30.
 XX
 XX Clostridium botulinum serotype G strain 113/30.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /note= "N-terminal His tag"
 FT
 XX
 XX WO9808540-A1.
 PN
 XX
 XX 05-MAR-1998.
 PD

```

XX 28-AUG-1997; 97WO-US15394.
PF 28-AUG-1996; 96US-0704159.
PR 28-AUG-1996; 96US-0704159.
XX (OPHI-) OPHIDIAN PHARM INC.
PA
XX Thalley BS, Williams JA;
PI
XX WPI: 1998-230234/20.
DR N-PSDB; AAV30596.
XX
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX
XX Example 49; Page 376-378; 428pp; English.
XX
XX This is the amino acid sequence of the histidine-tagged C fragment
XX of Clostridium botulinum (113/30 strain) type G neurotoxin, encoded
XX by a DNA sequence (see AAV30596) in plasmid pETN15b. This vector
XX can be used to express BotG soluble C fragment in Escherichia
XX coli host cells, with the recombinant C fragment being purified on
XX an affinity column. The invention relates to recombinant proteins
XX derived from C. botulinum toxins, especially type B and type E
XX toxins. Methods are provided which allow for the isolation of
XX soluble recombinant proteins free of significant endotoxin
XX contamination. Preferred hosts for production of recombinant
XX proteins are E. coli, insect cells and yeast cells. The
XX recombinant toxins are used as immunogens for the production of
XX vaccines and antitoxins that are useful in the treatment of humans
XX and animals at risk of intoxication with clostridial toxin.
XX
XX Sequence 473 AA:
SQ
XX
XX Query Match 10.3%; Score 116; DB 19; Length 473;
XX Best Local Similarity 22.2%; Pred. No. 0.096; Mismatches 13;
XX Matches 56; Conservative 33; Indels 88; Gaps 13;
XX
XX 5 HHHHHGSDPKATATTSGRKSTNVVHKSEAGTSSVFYKGTGDMLEPETHYRWFNINN 64
XX | | | | | : : : : : | | : : | |
XX 6 hhhhhhsghl-----egrhmasma-----dlilqvf---nn 36
XX
XX 65 EKSYSKDTIKDQIQGGQDLSTLNTINVTGTHSNYSGSASATDFEKAPFGSKTI 120
XX | : : : : : | : : : : : | : : : : : |
XX 37 ysnissnallslsyrgyllidsygaltnvgsdvlfdnigqgklnse---nsn 92
XX
XX 121 TVDNKNTIDVTIPGGYGS-VNSFSIN-----KTIITNDOCKEF-----VNSQAW 166
XX | : : : : : | : : : : : | : : : : : |
XX 93 fahgskfvv-----ydsmdafsnlwtvtrpkynndiqtylqneycliscikndsgw 145
XX
XX 167 -YOEHGKE-----EVNGKS-----FNHTVHNI---NANAGIEGT 196
XX | : : : : : | : : : : : | : : : : : |
XX 146 kvsignrliwrlidvnaaksksifeysikndisdylnkwfistltlndlganlyngs 205
XX
XX 197 VKGELKVLKQDK 208
XX : : : : : |
XX 206 lksekilnldtr 217
XX
XX RESULT 9
XX AAY00241
XX ID AAY00241 standard; Protein; 688 AA.
XX
XX AAY00241;
XX
XX 20-APR-1999 (first entry)
XX
XX Enterococcus faecalis antigenic polypeptide fragment EF124.
XX
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX detection; attenuation; antigenic.
XX

```

```

XX OS Enterococcus faecalis.
XX
XX W09850554-A2.
XX
XX 12-NOV-1998.
XX
XX 04-MAY-1998; 98WO-US08959.
XX
XX 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
XX 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bailey C, Choi GH, Bromocky] A, Kunsch CA;
XX
XX WPI: 1999-070095/06.
XX DR N-PSDB; AAX20231.
XX
XX New isolated Enterococcus faecalis polynucleotides - used to develop
XX products for the detection of Enterococcus and for use in vaccines
XX for prevention or attenuation of Enterococcus infection
XX
XX Claim 9; Page 230; 301pp; English.
XX
XX The present sequence represents an antigenic polypeptide fragment
XX isolated from Enterococcus faecalis. The present invention describes
XX genes, proteins and antigenic polypeptides isolated from E. faecalis.
XX The proteins can be used in vaccines for preventing or attenuating an
XX infection caused by a member of the Enterococcus genus in an animal.
XX They can also be used for detecting Enterococcus antibodies in a sample.
XX The nucleotide sequences can be used for detecting Enterococcus nucleic
XX acids. Products from the present invention can also be used for
XX screening compounds to identify agonists and antagonists of E. faecalis
XX protein activity.
XX
XX Sequence 688 AA:
SQ
XX
XX Query Match 10.2%; Score 114.5; DB 20; Length 688;
XX Best Local Similarity 25.5%; Pred. No. 0.22; Mismatches 93; Indels 29; Gaps 9;
XX Matches 54; Conservative 36; Indels 93; Indels 29; Gaps 9;
XX
XX 12 SDDKVAITISGNKSTNVVHKS-EAGTSSVFYKGTGDMLEPETHYRWFNINNKSYS 70
XX | : : : : : | : : : : : | : : : : : |
XX 172 sydnasylngssrdvtcgvsiqbhgsesv--kgeylnkddpdrhvyhvmingags-vl 228
XX
XX 71 KQITIKDQIQGGQDLSTLNTINVTGTHSNYSGSASATDFEKAPFGSKTIYDNKTN-TI 129
XX | : : : : : | : : : : : | : : : : : |
XX 229 ddvvltdlpspnqvldpeslvi--yftn-----vledgtlirpdksvllleegkycl 277
XX
XX 130 DVTIPGGYSYN-----SFSINXTKITNEQ--QKEFVNNSQAWYOEHGKEEVNG 177
XX | : : : : : | : : : : : | : : : : : |
XX 278 evrltneitgqgkivvkmahleapyymeyrslvtsaagstclvsn-qvasltngseevng 336
XX
XX 178 KSFNHTVININANAGIEGTVGKELKVLKQDK 209
XX | : : : : : |
XX 337 ddngdvvdvldhsgghatgkylkktamd 368
XX
XX RESULT 10
XX AAY00238
XX ID AAY00238 standard; Protein; 2032 AA.
XX
XX AAY00238;
XX
XX 20-APR-1999 (first entry)
XX
XX Enterococcus faecalis protein EF123.
XX
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX detection; attenuation; antigenic.
XX

```

```
XX OS Enterococcus faecalis.
XX PN W09850554-A2.
XX PD 12-NOV-1998.
XX PE 04-MAY-1998; 98MO-US08959.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;
XX WPI; 1999-070095/06.
XX DR N-PSDB; AAX20228.
XX PT New isolated Enterococcus faecalis polynucleotides - used to develop
XX PT products for the detection of Enterococcus and for use in vaccines
XX PT for prevention or attenuation of Enterococcus infection
XX PS Claim 9; Page 224-225; 301pp; English.
XX CC The present sequence represents a protein isolated from
XX CC Enterococcus faecalis. The present invention describes genes, proteins
XX CC and antigenic polypeptides isolated from E. faecalis. The proteins can
XX CC be used in vaccines for preventing or attenuating an infection caused
XX CC by a member of the Enterococcus genus in an animal. They can also be
XX CC used for detecting Enterococcus antibodies in a sample. The nucleotide
XX CC sequences can be used for detecting Enterococcus nucleic acids.
XX CC Products from the present invention can also be used for screening
XX CC compounds to identify agonists and antagonists of E. faecalis protein
XX CC activity.
XX SQ Sequence 2032 AA;

Query Match 10.2%; Score 114.5; DB 20; Length 2032;
Best Local Similarity 25.5%; Pred. No. 0.91;
Matches 54; Conservative 36; Mismatches 93; Indels 29; Gaps 9;

QY 12 SDDKVAATITSGNKSNTVTYHKS-EAGTSSVFFYRTGDMLPEDTTHVRFNLINNEKSYVS 70
DB 908 sydntlaasytngssrdvtygkvsighgsesv--kkgyeyhkddpdlhyvwhmlngags-vl 964
QY 71 KDITIKQIOGGQQLDLSTLININTGTHSNYSOGSALTDFEKAFFPSKTIYVDMTKN-TI 129
DB 965 ddvviltclpnpqyldpeslvi--ygtln-----vtedqtltpdksvllleegkdytl 1013
QY 130 DVTIPQGYGSYN-----SFSINKTKITNEQ--QKEFVNNSQAWYQEHGKEEYNG 177
DB 1014 evtcdnetggkqkiyvkmahleapyyrmejrslvtsaagstltvan-qvsltgngsevvhg 1072
QY 178 KSFNHTVHNINANAGIECTVKGELKVLKQDKD 209
DB 1073 ddngdvvvldhsgshgahgtgkylqkkttamd 1104

RESULT 11
AAY00240
ID AAY00240 standard; Protein; 2032 AA.
XX
XX AAY00240;
XX
XX 20-APR-1999 (first entry)
XX DE Enterococcus faecalis protein EPI24.
XX DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX KM detection; attenuation; antigenic.
```

```
XX OS Enterococcus faecalis.
XX PN W09850554-A2.
XX PD 12-NOV-1998.
XX PE 04-MAY-1998; 98MO-US08959.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;
XX WPI; 1999-070095/06.
XX DR N-PSDB; AAX20230.
XX PT New isolated Enterococcus faecalis polynucleotides - used to develop
XX PT products for the detection of Enterococcus and for use in vaccines
XX PT for prevention or attenuation of Enterococcus infection
XX PS Claim 9; Page 228-229; 301pp; English.
XX CC The present sequence represents a protein isolated from
XX CC Enterococcus faecalis. The present invention describes genes, proteins
XX CC and antigenic polypeptides isolated from E. faecalis. The proteins can
XX CC be used in vaccines for preventing or attenuating an infection caused
XX CC by a member of the Enterococcus genus in an animal. They can also be
XX CC used for detecting Enterococcus antibodies in a sample. The nucleotide
XX CC sequences can be used for detecting Enterococcus nucleic acids.
XX CC Products from the present invention can also be used for screening
XX CC compounds to identify agonists and antagonists of E. faecalis protein
XX CC activity.
XX SQ Sequence 2032 AA;

Query Match 10.2%; Score 114.5; DB 20; Length 2032;
Best Local Similarity 25.5%; Pred. No. 0.91;
Matches 54; Conservative 36; Mismatches 93; Indels 29; Gaps 9;

QY 12 SDDKVAATITSGNKSNTVTYHKS-EAGTSSVFFYRTGDMLPEDTTHVRFNLINNEKSYVS 70
DB 908 sydntlaasytngssrdvtygkvsighgsesv--kkgyeyhkddpdlhyvwhmlngags-vl 964
QY 71 KDITIKQIOGGQQLDLSTLININTGTHSNYSOGSALTDFEKAFFPSKTIYVDMTKN-TI 129
DB 965 ddvviltclpnpqyldpeslvi--ygtln-----vtedqtltpdksvllleegkdytl 1013
QY 130 DVTIPQGYGSYN-----SFSINKTKITNEQ--QKEFVNNSQAWYQEHGKEEYNG 177
DB 1014 evtcdnetggkqkiyvkmahleapyyrmejrslvtsaagstltvan-qvsltgngsevvhg 1072
QY 178 KSFNHTVHNINANAGIECTVKGELKVLKQDKD 209
DB 1073 ddngdvvvldhsgshgahgtgkylqkkttamd 1104

RESULT 12
AAY00242
ID AAY00242 standard; Protein; 2032 AA.
XX
XX AAY00242;
XX
XX 20-APR-1999 (first entry)
XX DE Enterococcus faecalis protein EPI25.
XX DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX KM detection; attenuation; antigenic.
```



```

Db      1 mrgshhhhhgscqalgqdmvspcatnsssfsspsagrhvrsyhlbgd---vrvrk 57
QY      61 NINNEKSYVSKDTITKDQIOGSGQ-----LDLSTLNINVTG---HSWYSGQSAITD 110
Db      58 lsfctylfll--lekngkvsqtkkencpyslltstvelgvvavkalsnyy---lamnk 112
QY      111 FEAFPGSKRTVD-NRKNTIDVTIPQGYGSYNSFSINYKTKRTNEQKKEPVNNSQAWYOE 169
Db      113 kgklylskqefndcklkerle---engyntyaafn-----wqn 147
QY      170 HGKE---EVNGK 178
Db      148 ngrqmyvalngk 159

RESULT 14
AAB10309
ID      AAB10309 standard; Protein; 184 AA.
XX
XX      AAB10309;
XX
XX      20-NOV-2000 (first entry)
XX
DE      Human KGF-2 construct pQE60-Cys37 protein SEQ ID NO: 30.
XX
KW      Human; keratinocyte growth factor; KGF-2; antiulcer; antidiabetic;
KW      antiinflammatory; cytoprotective; dermatological; gastrointestinal;
KW      hepatic; respiratory; renal; cerebroprotective; mucositis; treatment;
KW      epithelial cell proliferation; inflammatory bowel disease; lung damage;
KW      liver disorder; diabetes; oral injury; gastrointestinal injury;
KW      gut toxicity; gastric; duodenal; epidermolysis bullosa; skin graft;
KW      skin disorder; renal failure; brain injury; intestinal fibrosis;
KW      proctitis; female reproductive tract disorder; pulmonary fibrosis;
KW      pneumonitis; pleural retraction; hemopoietic syndrome; myelotoxicity.
XX
XX      Homo sapiens.
XX
OS      US6077692-A.
XX
PN      20-JUN-2000.
XX
PD      13-FEB-1998; 98US-0023082.
XX
PF      13-AUG-1996; 96US-0023852.
XX
PR      28-FEB-1997; 97US-0039045.
XX
PR      13-AUG-1997; 97US-0055561.
XX
PR      05-JUN-1995; 95US-0461195.
XX
PR      14-FEB-1995; 95WO-US01790.
XX
PR      23-MAY-1997; 97US-0862432.
XX
PR      13-AUG-1997; 97US-0910875.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Mendrick D, Duan DR, Ni J, Jimenez P, Coleman TA, Gruber JR.
PI      Dillon PJ, Gentz RL, Ruben SM, Zhang J, Moore PA, Rampy MA;
XX
DR      WPI: 2000-441307/38.
DR      N-PSDB: AAA71215.
XX
PT      Novel keratinocyte growth factor useful for promoting and accelerating
PT      wound healing, comprising at least 10 contiguous amino acids from a
XX
XX      specific amino acid sequence -
XX
XX      Example 7; Fig 15; 190pp; English.
XX
XX      This invention describes a novel human keratinocyte growth factor, KGF-2
XX      (I), which has antiulcer, antidiabetic, antiinflammatory, cytoprotective,
XX      dermatological, gastrointestinal, hepatic, respiratory, renal and
XX      cerebroprotective activity. (I) is useful for stimulating epithelial cell
XX      proliferation in patients suffering from wound, mucositis, ulcer such as
XX      venous stasis ulcer, diabetic ulcer and cubitus ulcer. (I) is also useful
XX      for treating inflammatory bowel disease, liver disorder, lung damage,
XX      diabetes, oral injury, gastrointestinal injury, gut toxicity, gastric

```

```

CC      ulcer, duodenal ulcer, epidermolysis bullosa, skin graft, skin disorder,
CC      renal failure, brain injury, breast tissue injury, urothelial damage,
CC      female reproductive tract disorder, intestinal fibrosis, proctitis,
CC      pulmonary fibrosis, pneumonitis, pleural retraction, hemopoietic syndrome
CC      and myelotoxicity. (I) is also useful for increasing the adherence of
CC      skin grafts to wound beds and to stimulate re-epithelialization from the
CC      wound bed, to produce changes in hepatocyte proliferation, to reduce the
CC      side effects of gut toxicity, to regenerate skin in full and partial
CC      thickness skin defects, and to prevent and heal damage to lungs. KGF-2
CC      shows enhanced activity, increased stability, higher yield and better
CC      solubility. This sequence represents the human KGF-2 protein construct
XX      pQE60-Cys37 described in the method of the invention.
XX
SQ      Sequence 184 AA;

Query Match      10.1%; Score 113.5; DB 21; Length 184;
Best Local Similarity 23.4%; Pred. No. 0.045;
Matches 45; Conservative 35; Mismatches 65; Indels 47; Gaps 9;

QY      1 MRSHHHHHSDKVAATITSGNKSITNVYKSEAGTSYFYKTDLPDTHVWML 60
Db      1 mrgshhhhhgscqalgqdmvspcatnsssfsspsagrhvrsyhlbgd---vrvrk 57
QY      61 NINNEKSYVSKDTITKDQIOGSGQ-----LDLSTLNINVTG---HSWYSGQSAITD 110
Db      58 lsfctylfll--lekngkvsqtkkencpyslltstvelgvvavkalsnyy---lamnk 112
QY      111 FEAFPGSKRTVD-NRKNTIDVTIPQGYGSYNSFSINYKTKRTNEQKKEPVNNSQAWYOE 169
Db      113 kgklylskqefndcklkerle---engyntyaafn-----wqn 147
QY      170 HGKE---EVNGK 178
Db      148 ngrqmyvalngk 159

RESULT 15
AAB61673
ID      AAB61673 standard; Protein; 184 AA.
XX
XX      AAB61673;
XX
XX      10-APR-2001 (first entry)
XX
DE      pQE60-Cys37 construct protein.
XX
KW      Keratinocyte growth factor; KGF-2; epithelial cell proliferation; wound;
KW      mucositis; ulcer; inflammatory bowel disease; liver disorder;
KW      lung damage; diabetes; oral injury; gastrointestinal injury;
KW      epidermolysis bullosa; renal failure; brain injury; proctitis;
KW      pulmonary fibrosis; haemopoietic syndrome; ovary injury; infertility;
XX      liver fibrosis.
XX
XX      Unidentified.
XX
OS      WO200102433-A1.
XX
PN      11-JAN-2001.
XX
PD      03-JUL-2000; 2000WO-US18328.
XX
PF      02-JUL-1999; 99US-0142343.
XX
PR      14-JUL-1999; 99US-0143648.
XX
PR      15-JUL-1999; 99US-0144024.
XX
PR      12-AUG-1999; 99US-0148628.
XX
PR      19-AUG-1999; 99US-0149935.
XX
PR      03-NOV-1999; 99US-0163375.
XX
PR      22-DEC-1999; 99US-0171677.
XX
PR      19-APR-2000; 2000US-0198322.
XX
PR      19-MAY-2000; 2000US-0205417.
XX
PR      30-JUN-2000; 2000US-9911224.
XX

```

PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Jimenez P, Duan DR, Rampy MA, Mendrick D, Zhang J;
PI Ni J, Moore PA, Coleman TA, Gruber JR, Dillon PJ, Gentz RL;
XX WPI: 2001-071578/08.
DR N-PSDB: AAF31913.

XX A polynucleotide encoding the human keratinocyte growth factor useful
PT for stimulating epithelial cell proliferation in a patients that has
PT e.g a wound -

XX Example 8; Fig 15; 591pp; English.

XX The present invention relates to human keratinocyte growth factor (KGF-2;
CC see AAF31901 and AAF61657). The present sequence is a KGF-2 construct
CC protein, which was generated in the present invention. KGF-2 can be used
CC to stimulate epithelial cell proliferation in a patient, where the
CC patient has a wound, mucositis, an ulcer, inflammatory bowel disease,
CC liver disorder, lung damage, diabetes, oral injury, gastrointestinal
CC injury, gut toxicity, epidermolysis bullosa, skin graft, skin disorder,
CC renal failure, brain injury, breast tissue injury, urothelial damage,
CC female reproductive tract disorder, intestinal fibrosis, proctitis,
CC pulmonary fibrosis, penumonia, plural retraction, haemopoietic
CC syndrome, and myelotoxicity. In addition, KGF-2 can be used in the
CC treatment or prevention of ovary injury, infertility, or fibrosis of the
CC liver. KGF-2 also promotes internal healing, donor site healing, internal
CC surgical wound healing or healing of incisional wounds made during
CC cosmetic surgery in a patient.

XX Sequence 184 AA;

Query Match 10.1%; Score 113.5; DB 22; Length 184;

Best Local Similarity 23.4%; Pred. No. 0.045; Mismatches 65; Indels 47; Gaps 9;

Matches 45; Conservative 35; Mismatches 65; Indels 47; Gaps 9;

OY 1 MGSNNHHHSGDDVATTSNGKSTNTVHKSEAGTSVFYKGTGDMLEPDTTHVRMFL 60
DB . 1 mrgshhnhhgscqalqgdmvgspeatnsssfsspsagrhvrsynhlqgd---vvrkx 57
OY 61 NNNNEKSYVSKDITTKDQIGGQ-----LDLSTLNINVTGT--HSNYSGSQAITD 110
DB 58 lfstkkytlk--lekngvsgtkencpyslletlsvelgvavakainsny---lamnk 112
OY 111 FEKAPPGSKITVD-NTKNTIDVTIPQGYGSYNSFSINVKTKITNEQKFEVNNQAWQGE 169
DB 113 kgklyskfeindcklkerie---engyntlyastn-----wqh 147
OY 170 HGKE--EVNGK 178
DB 148 ngrqmyvalngk 159

Search completed: January 30, 2002, 00:01:07
Job time: 4595 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 23:13:57 ; Search time 65.03 Seconds
(without alignments)
73.015 Million cell updates/sec

Title: US-09-813-820-4

Sequence: 1 MRGSHHHHHSDSKVATIT.....GIEGVKGLKVLKODKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1122	100.0	211	4	US-08-856-253-4
2	1054.5	94.0	512	4	US-08-856-253-6
3	1048	93.4	1183	2	US-08-447-031A-2
4	830	74.0	159	4	US-08-856-253-2
5	125.5	11.2	345	4	US-08-856-253-7
6	113.5	10.1	184	3	US-09-023-082A-30
7	110.5	9.8	1338	4	US-08-728-470-9
8	110.5	9.8	1338	4	US-08-719-641-9
9	110.5	9.8	1599	2	US-08-617-697-9
10	104.5	9.3	1529	2	US-08-728-470-10
11	104.5	9.3	1529	4	US-08-719-641-10
12	104.5	9.3	1600	2	US-08-617-697-10
13	103.5	9.2	139	4	US-08-856-253-8
14	100	8.9	1536	1	US-08-038-682-2
15	100	8.9	1536	1	US-08-302-832-2
16	100	8.9	1536	2	US-08-530-198-2
17	100	8.9	1536	2	US-08-469-880-2
18	100	8.9	1536	2	US-08-728-470-2
19	100	8.9	1536	2	US-08-617-697-2
20	100	8.9	1536	4	US-08-719-641-2
21	99.5	8.8	589	4	US-09-377-155-19
22	98.5	8.8	351	2	US-08-701-191A-3
23	95	8.5	666	2	US-08-737-716-14
24	91	8.1	599	1	US-08-172-331B-4
25	90	8.0	76	3	US-08-956-307B-17
26	90	8.0	76	3	US-08-956-307B-18
27	90	8.0	122	3	US-08-956-307B-19

28	90	8.0	2123	4	US-08-968-685A-10	Sequence 10, Appl
29	89.5	8.0	1702	5	US-08-296-791-5	Sequence 5, Appl
30	89.5	8.0	1702	5	PCT-US85-10661A-5	Sequence 5, Appl
31	89	7.9	599	2	US-09-032-315-6	Sequence 6, Appl
32	89	7.9	599	2	US-08-993-318A-6	Sequence 6, Appl
33	89	7.9	599	4	US-09-399-886-6	Sequence 6, Appl
34	89	7.9	599	4	US-09-396-260-6	Sequence 6, Appl
35	89	7.9	599	4	US-09-576-281-6	Sequence 6, Appl
36	89	7.9	631	1	US-08-605-541B-12	Sequence 12, Appl
37	87	7.8	254	4	US-09-004-731-89	Sequence 92, Appl
38	87	7.8	254	4	US-09-004-731-92	Sequence 89, Appl
39	87	7.8	254	4	US-08-749-699-89	Sequence 89, Appl
40	87	7.8	254	4	US-08-749-699-92	Sequence 92, Appl
41	86.5	7.7	2353	4	US-09-377-155-33	Sequence 33, Appl
42	86.5	7.7	2353	4	US-08-913-942-4	Sequence 4, Appl
43	86	7.7	216	2	US-08-737-716-4	Sequence 4, Appl
44	86	7.7	310	3	US-09-136-628-2	Sequence 2, Appl
45	86	7.7	437	2	US-08-737-716-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-856-253-4
Sequence 4, Application US/08856253
Patent No. 6288214
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Patti, Joseph M.
APPLICANT: House-Pompeo, Karen
APPLICANT: Stahanam, Narayana
APPLICANT: Symersky, Jindrich
TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,253
CLASSIFICATION: 514
FILING DATE: Concurrently Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TAMK:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 211 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-856-253-4
Query Match 100.0%; Score 1122; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.1e-94;

Db 1 MRGSHHHHGSVNAADAPAGTDITNLTNTVIGIDSGTTVVPHOAGYKLVN---GFS 57
 QY 49 LPEDTTHVHMF-----LNINNEKSYVS-KDITIKDOLOGGOQLDLSTLNTINVTGHSN 100
 Db 58 VPRSAVKGDFEKTIVKELNNGVSTAKVPRIMAGDOVLANSVIDSDG---NVITTFD 114
 QY 101 YVSGQSAITDFEKA---FPGSKITVDNTKNTIDVTIPOGYGSYNSFSINXKTKITNEQOK 157
 Db 115 YVWTK----DDVYKATLTPM-AYIDPENVKKTGNVTLATIGISTTA---KKTVDYDEKYG 166
 QY 158 EFPVNS 163
 Db 167 KEYNLS 172

RESULT 6
 US-09-023-082A-30
 ; Sequence 30, Application US/09023082A
 ; Patent No. 6077692
 ; GENERAL INFORMATION:
 ; APPLICANT: RUBEN, STEVEN M.
 ; APPLICANT: JIMENEZ, PABLO
 ; APPLICANT: DUAN, D. ROXANNE
 ; APPLICANT: RAMPY, MARK A.
 ; APPLICANT: MENDRICK, DONNA
 ; APPLICANT: ZHANG, JUN
 ; APPLICANT: NI, JIAN
 ; APPLICANT: MOORE, PAUL A.
 ; APPLICANT: COLEMAN, TIMOTHY A.
 ; APPLICANT: GRUBER, JOACHIM R.
 ; APPLICANT: DILLON, PATRICK J.
 ; APPLICANT: GENTZ, REINER L.
 ; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
 ; STREET: 1100 NEW YORK AVE, NW, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/023,082A
 ; FILING DATE: 13-FEB-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/011790
 ; FILING DATE: 14-FEB-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/461,195
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/023,852
 ; FILING DATE: 13-AUG-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/039,045
 ; FILING DATE: 28-FEB-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/862,432
 ; FILING DATE: 23-MAY-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/910,875
 ; FILING DATE: 13-AUG-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/055,561
 ; FILING DATE: 13-AUG-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-2600
 TELEFAX: 202-371-2540
 ; INFORMATION FOR SEQ ID NO: 30:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 184 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-023-082A-30

Query Match 10.1%; Score 113.5; DB 3; Length 184;
 Best Local Similarity 23.4%; Pred. No. 0.0046;
 Matches 45; Conservative 35; Mismatches 65; Indels 47; Gaps 9;

QY 1 MRGSHHHHGSDDKVAATTSCKNSTVTVHKSAGTSSVFYKGTGDMLPEDTTHVHMF 60
 Db 1 MRGSHHHHGSQALGQDWVSPDATNSSSSPSPSAGRHVRSYNHLOGD---VRWRK 57
 QY 61 NINNEKSYVSKDITTDQIOGGQ-----LDLSTLNTINVTG---HSNYSGQSAITD 110
 Db 58 LFSFTYFLK--LEKNGKVSQTKREKCPYSILETTSVEIGVAVKAINSYX---LAWRK 112
 QY 111 FEKAFPGSKITVD-NTRKNIDVTIPOGYGSYNSFSINXKTKITNEQOKFEVNSQAWOE 169
 Db 113 KGLYSGKEFNNDCKIKERE---ENGINTYASFN-----MQH 147
 QY 170 HGKE---EVNGK 178
 Db 148 NGRQMYVALNGK 159

RESULT 7
 US-08-728-470-9
 ; Sequence 9, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,470
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 9.8%; Score 110.5; DB 2; Length 1338;
Best Local Similarity 25.5%; Pred. No. 0.14;
Matches 61; Conservative 28; Mismatches 89; Indels 61; Gaps 10;

QY 12 SDDKVAITISGNKST-----NTVHKSEA-----GTSVFEYKTDMLPEDTHVWFNLN 61
DB 694 SDTYRTITIKGNISNKGDLNIIIDKSDAEIIGNIS---QKEGN-----LT 738
QY 62 INNEKSVSKDITIKDQIGGQQLDSTLNINVT-GTHSNYSGSAITDFEKAFFPSKI 120
DB 739 ISSDKVNIITQITIKAGVEGGRSDSEAEANLTIQTKEKLKLAGDLNIGFNKA----EI 794
QY 121 TVDNKNTIDVTIPQGYGSI-----NSFSINVKTKITNEQKEFY 160
DB 795 T---AKNGSDLTIGNAGGNADAKKVTEDKVKDSKISTDGHNVTLNSEVKTSGSSNAGN 851
QY 161 NNSQAMQOEHGKEEVNGKSFNHTVHNINANAG----IEGT----VKELKVLKQDKDTK 211
DB 852 DNSTGLTISAKDVTYNNVNTSHKTTINISAAAGVTTKEGTTINATTGSVEYTAQNGTIK 910

RESULT 8

US-08-719-641-9
Sequence 9, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-9

Query Match 9.8%; Score 110.5; DB 4; Length 1338;
Best Local Similarity 25.5%; Pred. No. 0.14;
Matches 61; Conservative 28; Mismatches 89; Indels 61; Gaps 10;

QY 12 SDDKVAITISGNKST-----NTVHKSEA-----GTSVFEYKTDMLPEDTHVWFNLN 61
DB 694 SDTYRTITIKGNISNKGDLNIIIDKSDAEIIGNIS---QKEGN-----LT 738
QY 62 INNEKSVSKDITIKDQIGGQQLDSTLNINVT-GTHSNYSGSAITDFEKAFFPSKI 120
DB 739 ISSDKVNIITQITIKAGVEGGRSDSEAEANLTIQTKEKLKLAGDLNIGFNKA----EI 794
QY 121 TVDNKNTIDVTIPQGYGSI-----NSFSINVKTKITNEQKEFY 160
DB 795 T---AKNGSDLTIGNAGGNADAKKVTEDKVKDSKISTDGHNVTLNSEVKTSGSSNAGN 851
QY 161 NNSQAMQOEHGKEEVNGKSFNHTVHNINANAG----IEGT----VKELKVLKQDKDTK 211
DB 852 DNSTGLTISAKDVTYNNVNTSHKTTINISAAAGVTTKEGTTINATTGSVEYTAQNGTIK 910

RESULT 9

US-08-617-697-9
Sequence 9, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9

Query Match 9.8%; Score 110.5; DB 2; Length 1599;
Best Local Similarity 25.5%; Pred. No. 0.18;
Matches 61; Conservative 28; Mismatches 89; Indels 61; Gaps 10;

QY 12 SDDKAVTISGKST-----NTYHKSSEA-----GTSVFFYKTCGMLPEDTTHRWFL 61
DB 954 SPTTYRTIKKINSKSGDLITDKDAEIQIGNIS---QKEG-----LT 998
QY 62 INNERSYVSKDITKIDQOGGOOLDSTLINVT--GTHSNYSGSATIDFEKAPGSK 120
DB 999 ISSDKYNTINQTTIKAGVGRSDSEAEANALTIQTKELKLAGDLINISGFNKA-----E 1054
QY 121 TVDNTKNTIDVTIPOGYGSY-----NSFSINYKTKITNEQOKEFY 160
DB 1055 T---AKNSGDLITIGNASGGMADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAG 1111
QY 161 NNSQAMYGHEKKEEVNGKSFNHTVHINANAG-----IEGT---VKGLKVLKQDKDTK 211
DB 1112 DNSTGLTISAKDVTYNNNVTSKRTINISAAAGNVTTKEGTTINATGTSVEYAQNGTIR 1170

RESULT 10

US-08-728-470-10

; Sequence 10, Application US/08728470

; Patent No. 5928651

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/728,470

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 16-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9205704.1

; FILING DATE: 16-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstresser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-633

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 9.3%; Score 104.5; DB 2; Length 1529;
Best Local Similarity 22.9%; Pred. No. 0.58;
Matches 55; Conservative 32; Mismatches 90; Indels 63; Gaps 9;

QY 16 VATITSGKST-----NTYHKSSEA-----FYKTDMLPEDTTHRWFL 60
DB 881 ITTNAGTQRTIINGNITKEKGLNLIKADAEIQIGNISQKEN-----L 928
QY 61 INNERSYVSKDITKIDQOGGOOLDSTLINVT--GTHSNYSGSATIDFEKAPGSK 119
DB 929 TISSDKYNTINQTTIKAGVGRSDSEAEANALTIQTKELKLAGDLINISGFNKA-----E 984
QY 120 TVDNTKNTIDVTIPOGYGSY-----NSFSINYKTKITNEQOKEFY 159
DB 985 IT---AKNSGDLITIGNASGGMADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAG 1041
QY 160 VNSQAMYGHEKKEEVNGKSFNHTVHINANAG-----IEGT---VKGLKVLKQDKDTK 211
DB 1042 DNSTGLTISAKDVTYNNNVTSKRTINISAAAGNVTTKEGTTINATGTSVEYAQNGTIR 1101

RESULT 11

US-08-719-641-10

; Sequence 10, Application US/08719641

; Patent No. 6218141

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/719,641

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 16-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9205704.1

; FILING DATE: 16-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstresser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-625

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1529 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 9.3%; Score 104.5; DB 4; Length 1529;
Best Local Similarity 22.9%; Pred. No. 0.58;
Matches 55; Conservative 32; Mismatches 90; Indels 63; Gaps 9;

QY 16 VAITSGNKST---NVTVHKSSEAGTSSV-----FYKKGMDLPEDTHVWFL 60
DB 881 ITTNASGTOQTIIINGNTINEKGLNKNIKADAEIQIGINSQKEGN-----L 928
QY 61 NINNESYVSKDITIKDQIQGOOLDLSTLINVT-GTHSNYSQGSATIDFEKAPGSK 119
DB 929 TISSDKVNITNOITIKAGVGRSDSSEANMLTIOTKELKLAGDLNISGFNKA-----E 984
QY 120 ITVDNKNITIDVTIPOGYSY-----NSFSINYKTKITNEOKEF 159
DB 985 IT---AKNSGDLTIGNASGGMADAKKVTPEKVKDSKISTDGHNVTLNSEVKTSSNGSSNAG 1041
QY 160 VNNSQAMYOEHGKEEYNGKSFNHTVHNINANAG---IEGT---VKGELKVLKODKDK 211
DB 1042 NDNSTGLTISAKDVTYNNVNTSHKTTINISAAAGVTTKEGTITNATGSEVYTAQNGTIK 1101

RESULT 12
US-08-617-697-10
Sequence 10, Application US/08617697
Patent No. 5977336

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1

CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

Query Match 9.3%; Score 104.5; DB 2; Length 1600;
Best Local Similarity 22.9%; Pred. No. 0.62;
Matches 55; Conservative 32; Mismatches 90; Indels 63; Gaps 9;

QY 16 VAITSGNKST---NVTVHKSSEAGTSSV-----FYKKGMDLPEDTHVWFL 60
DB 951 ITTNASGTOQTIIINGNTINEKGLNKNIKADAEIQIGINSQKEGN-----L 998
QY 61 NINNESYVSKDITIKDQIQGOOLDLSTLINVT-GTHSNYSQGSATIDFEKAPGSK 119
DB 999 TISSDKVNITNOITIKAGVGRSDSSEANMLTIOTKELKLAGDLNISGFNKA-----E 1054
QY 120 ITVDNKNITIDVTIPOGYSY-----NSFSINYKTKITNEOKEF 159
DB 1055 IT---AKNSGDLTIGNASGGMADAKKVTPEKVKDSKISTDGHNVTLNSEVKTSSNGSSNAG 1111
QY 160 VNNSQAMYOEHGKEEYNGKSFNHTVHNINANAG---IEGT---VKGELKVLKODKDK 211
DB 1112 NDNSTGLTISAKDVTYNNVNTSHKTTINISAAAGVTTKEGTITNATGSEVYTAQNGTIK 1171

RESULT 13
US-08-856-253-8
Sequence 8, Application US/08856253
Patent No. 6288214

GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Patti, Joseph M.
APPLICANT: House-Pompeo, Karen
APPLICANT: Stnam, Narayana
APPLICANT: Symersky, Jindrich
TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston

STATE: Texas
COUNTRY: U.S.
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/856,253
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996

ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TAMK:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-856-253-8

Query Match 9.2%; Score 103.5; DB 4; Length 139;
Best Local Similarity 25.7%; Pred. No. 0.025;

Matches 46; Conservative 16; Mismatches 54; Indels 63; Gaps 9;

OY 1 MRSHHHHHSDDKVATTSNGKSTNTVHKSEAGTSVFYTKGDM-LPDTTHRMFL 60
 |||||
 Db 1 MRSHHHHHSE-----GGQNSGN-----FEEDTE----- 28

OY 61 NINNEKSYVSKITITKDOIGGQQLDLSTLINVT-GTHSNYSGSATIDFEKAPGSKI 120
 ::|
 Db 29 --EDKPKYE-----QSGNIVDIDF--DSVPOIHQONKGNOSFEEDTEK---DKP 70

OY 121 TPDNKTNTIDV--TIPOGYSYNSFSINYTKTKITEQOKREFVNNNSOAWYOEHEKEEVN 176
 ::|||
 Db 71 KKHGHNIIIDIDFDSVPHIHG-----FKKHTEIIEED---TNKDKPSYOGGHNHNSVD 119

RESULT 14
 US-08-038-682-2
 ; Sequence 2, Application US/08038682
 ; Patent No. 5549897
 ; GENERAL INFORMATION:
 ; APPLICANT: BARENKAMP, STEPHEN J
 ; APPLICANT: ST. GEME III, JOSEPH W
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
 ; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Maltare, Ltd
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/038, 682
 ; FILING DATE: 16-MAR-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BERKSTRESSER, JERRY W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-293
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1536 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-038-682-2

Query Match 8.9%; Score 100; DB 1; Length 1536;
 Best local similarity 25.2%; Pred. No. 1.5;
 Matches 51; Conservative 33; Mismatches 72; Indels 46; Gaps 9;

Db 1057 DGSDLITGNT-NSADGT-----NAKKVTENQVADSKISADGHVTL-----HS 1098

OY 172 KEFVNGKSFNHTVHNINANAGI 193
 |::|
 Db 1099 KVETSG-SNNNTEDSSDNNAGL 1119

RESULT 15
 US-08-302-832-2
 ; Sequence 2, Application US/08302832
 ; Patent No. 5603938
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Maltare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/302, 832
 ; FILING DATE: 16-SEP-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US pct/us93/02166
 ; FILING DATE: 16-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-404
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1536 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-302-832-2

Query Match 8.9%; Score 100; DB 1; Length 1536;
 Best local similarity 25.2%; Pred. No. 1.5;
 Matches 51; Conservative 33; Mismatches 72; Indels 46; Gaps 9;

OY 13 DDKVATITSGKSTNTVHKSEAGTSVFYTKGDM-----LPDTTHRMFL 60
 :|::|
 Db 943 NSKNLSITNTSSSTYRTI-----ISGNITNKNGDLNITNEGSDTEMQIGGDVSQREGNL 996

OY 61 NINNEKSYVSKITITKDOIGGQQLDLSTLINVT-GTHSNYSGSATIDFEKAPGSKI 115
 |::|
 Db 997 TISSDKINIKQITITAGVDGNSDSDATNNANLITIKTELKLTQDLNLSGFKAEITAK 1056

OY 116 PGSKITVDNKTNTIDVITPOGYSYNSFSINX-----TKITNEQOKREFVNNNSOAWYOEHG 171
 |::|
 Db 1057 DGSDLITGNT-NSADGT-----NAKKVTENQVADSKISADGHVTL-----HS 1098

OY 172 KEFVNGKSFNHTVHNINANAGI 193

Thu Jan 31 10:59:13 2002

us-09-813-820-4.rail

Page 9

Db 1099 KVEISG-SNNTEDESSDNNAGL 1119

Search completed: January 30, 2002, 00:02:24
Job time: 2907 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 00:02:27 ; Search time 49.08 Seconds

(without alignments)
157.626 Million cell updates/sec

Title: US-09-813-820-4

Perfect score: 1122
Sequence: 1 MRGSHHHHHGSDKVAFT.....GIEGVKGLKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048	93.4	1183	1 CNA_STAU	Q53654 staphylococ
2	106	93.4	1076	1 HSER_CAVPO	P70106 cavia porce
3	98	87.7	1256	1 ATL_STAU	P52081 staphylococ
4	96.5	8.6	848	1 TOP1_BORBU	O51768 borrelia bu
5	96	8.6	704	1 MSN2_YEAST	P33748 saccharomyc
6	95	8.5	666	1 MUR2_ENTHR	P39046 enterococcu
7	95	8.5	2003	1 YDBA_ECOLI	P33666 escherichia
8	93.5	8.3	1296	1 BXG_CLOBO	P33666 escherichia
9	92.5	8.2	711	1 TOP1_MYCPN	P78032 mycoplasma
10	91.5	8.2	238	1 OPAH_NERGO	O04884 neisseria g
11	91	8.1	463	1 GUN_BACSP	P29019 bacillus sp
12	91	8.1	578	1 LIPA_BACPU	Q50774 mycoplasma
13	90.5	8.1	1169	1 CIGB_BACTZ	O92426 bacillus th
14	89.5	8.0	416	1 ELT1_CAEEL	P28515 caenorhabdi
15	89	7.9	599	1 LAC2_THACU	O02075 tharaphor
16	89	7.9	858	1 CYAG_DICDI	O03101 dictyostell
17	88.5	7.9	520	1 AMT_BACME	P20845 bacillus me
18	88.5	7.9	717	1 UBPB_YEAST	P36026 saccharomyc
19	87.5	7.8	382	1 Y077_METUA	Q60384 methanococc
20	87.5	7.8	523	1 RPN3_YEAST	P40016 saccharomyc
21	87.5	7.8	626	1 PC11_YEAST	P39681 saccharomyc
22	87.5	7.8	1251	1 RBR2_PLAVB	O00799 plasmodium
23	87.5	7.8	1435	1 DPO3_MYCPU	P47729 mycoplasma
24	87	7.8	668	1 PBPC_BACPU	P42971 bacillus su
25	87	7.8	762	1 SLAP_ACEKI	P22358 acetogenium
26	87	7.8	1433	1 SUBE_BACSU	P16397 bacillus su
27	86.5	7.7	475	1 Y123_MYCPN	P75513 mycoplasma
28	86.5	7.7	678	1 YNC7_YEAST	P53368 saccharomyc
29	86.5	7.7	1694	1 IGA0_HAEIN	P44569 haemophilus
30	86.5	7.7	1702	1 IGA2_HAEIN	P45884 haemophilus
31	86	7.7	437	1 ACMA_LACLC	O48603 lactococcus
32	86	7.7	617	1 Y237_BUCAT	P57331 buchnera ap
33	86	7.7	986	1 GUNZ_CLOSR	P23859 clostridium

34	85.5	7.6	938	1 PM15_CHLPN	O92883 chlamydia p
35	85.5	7.6	1146	1 AS10_YEAST	P48361 saccharomyc
36	85.5	7.6	1277	1 IF3X_YEAST	O03690 saccharomyc
37	85	7.6	691	1 P21_YEAST	P26570 saccharomyc
38	85	7.6	1088	1 VATA_CANTR	P38078 candida tiro
39	84.5	7.5	415	1 EP1G_YEAST	P29547 saccharomyc
40	84.5	7.5	925	1 PDC2_YEAST	P32896 saccharomyc
41	84.5	7.5	1569	1 YPJA_ECOLI	P52143 escherichia
42	84.5	7.5	1682	1 MSP1_PLAF3	P19598 plasmodium
43	84.5	7.5	2652	1 RRPB_IBVB	P26314 avian infec
44	84	7.5	401	1 FLGE_ECOLI	P75937 escherichia
45	84	7.5	437	1 YABE_BACSU	P37546 bacillus su

ALIGNMENTS

```

RESULT 1
ID      CNA_STAU          STANDARD;          PRT; 1183 AA.
AC      053654;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DE      15-DEC-1998 (Rel. 37, Last annotation update)
GN      COLLAGEN ADHESIN PRECURSOR.
OS      CNA.
OC      Staphylococcus aureus.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Staphylococcus.
OX      NCBI_TaxID=1280;
RN      [1]
RP      SPOUNCE FROM N.A.
RC      STRAIN-FDA 574;
RX      MEDLINE=92165839; PubMed=1311320;
RA      Patil J.M., Jonsson H., Guss B., Switalski L.M., Wlberg K.,
RA      Lindberg M., Hoeek M.;
RT      "Molecular characterization and expression of a gene encoding a
RT      Staphylococcus aureus collagen adhesin.";
RL      J. Biol. Chem. 267:4766-4772(1992).
RN      [2]
RP      ERRATUM.
RA      Patil J.M., Jonsson H., Guss B., Switalski L.M., Wlberg K.,
RA      Lindberg M., Hoeek M.;
RL      J. Biol. Chem. 269:11672-11672(1994).
RN      [3]
RP      COLLAGEN-BINDING DOMAIN.
RC      STRAIN-FDA 574;
RX      MEDLINE=9403261; PubMed=8218209;
RA      Patil J.M., Boles J.O., Hoeek M.;
RT      "Identification and biochemical characterization of the ligand
RT      binding domain of the collagen adhesin from Staphylococcus aureus.";
RL      Biochemistry 32:11428-11435(1993).
RN      [4]
RP      X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX      MEDLINE=97475225; PubMed=9334749;
RA      Smyersky J., Patil J.M., Carson M., House-Pompeo K., Teale M.,
RA      Moore D., Jin L., Schneider A., Delucas L.J., Hoeek M.,
RA      Narayana S.V.L.;
RT      "Structure of the collagen-binding domain from a Staphylococcus
RT      aureus adhesin.";
RL      Nat. Struct. Biol. 4:833-838(1997).
CC      -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCCAL CELLS TO
CC      COLLAGEN-CONTAINING SUBSTRATA.
CC      -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL WALL.
CC      -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCCAL PROTEINS
CC      IN THE REGION OF THE MEMBRANE ANCHOR.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch.

 CC EMBL; M81736; AAA20874.1; .
 DR PDB; 1AMX; 24-JUN-98.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 1183
 FT DOMAIN 30 1157
 FT TRANSMEM 1158 1177
 FT DOMAIN 1178 1183
 FT DOMAIN 151 318
 FT DOMAIN 533 1093
 FT DOMAIN 1093 1157
 FT DOMAIN 1151 1156
 FT REPEAT 533 719
 FT REPEAT 720 906
 FT REPEAT 907 1093
 FT REPEAT B1.
 FT REPEAT B2.
 FT REPEAT B3.
 SQ SEQUENCE 1183 AA; 133066 MW; 66A1CC072E575D76 CRC64;

Query Match 93.4%; Score 1048; DB 1; Length 1183;
 Best Local Similarity 100.0%; Pred. No. 1,le-69;
 Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

12 SDDKATITSGNKTSTVWTHVHKSSEAGTSVFYKKTGDMLEPDTTHVWFLINNEKSYVSK 71
 144 SDDKATITSGNKTSTVWTHVHKSSEAGTSVFYKKTGDMLEPDTTHVWFLINNEKSYVSK 203
 72 DITIKQIOGGQOLDLSTLININVTGTHSNVYSGQSAITDFEKAFGSKITVDNKTITIV 131
 204 DITIKQIOGGQOLDLSTLININVTGTHSNVYSGQSAITDFEKAFGSKITVDNKTITIV 263
 132 TTPQGGTSGNSINIKTKITNEQCKEFPVNSQAWQEGKKEVNSKSNHYHNINANA 191
 264 TTPQGGTSGNSINIKTKITNEQCKEFPVNSQAWQEGKKEVNSKSNHYHNINANA 323
 192 GIEGYKGLKYLKODKDK 211
 324 GIEGYKGLKYLKODKDK 343

RESULT 2
 HSER_CAVPO STANDARD; PRT: 1076 AA.
 ID HSER_CAVPO
 AC P70106;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL
 DE GUANYLYL CYCLASE) (EC 4.6.1.2) (STA RECEPTOR) (GUANYLYL CYCLASE C).
 GN GUCY2C OR GUCYC.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 OX NCBI_Taxid=10141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Kruhoefter M., Cetin Y., Kaempf U., Forssmann W.-G.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RECEPTOR FOR THE E. COLI HEAT-STABLE ENTEROTOXIN (E. COLI
 CC ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF GMP IN
 CC MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
 CC PEPTIDE GUANYLIN.
 CC -1- CATALYTIC ACTIVITY: GTP = 3, '5'-CYCLIC GMP + PYROPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; Z74734; CAA98989.1; .
 DR HSSP; Q02846; IAWL.
 DR InterPro; IPR001828; ANF_receptor.
 DR InterPro; IPR00719; Euk_pkinase.
 DR InterPro; IPR01054; Guanylyl_cyclase.
 DR Pfam; PF01094; ANF_receptor; 1.
 DR Pfam; PF00211; guanylate_cyc; 1.
 DR Pfam; PF00069; pkinase; 2.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLYL_CYCLASES_1; 1.
 DR PROSITE; PS00125; GUANYLYL_CYCLASES_2; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Receptor; Transmembrane; Glycoprotein; Lyase; GMP synthesis; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 1076
 FT DOMAIN 24 433
 FT TRANSMEM 434 457
 FT DOMAIN 458 1076
 FT DOMAIN 492 752
 FT DOMAIN 827 957
 FT CARBOHYD 35 35
 FT CARBOHYD 82 82
 FT CARBOHYD 191 191
 FT CARBOHYD 191 191
 FT CARBOHYD 198 198
 FT CARBOHYD 287 287
 FT CARBOHYD 306 306
 FT CARBOHYD 310 310
 FT CARBOHYD 348 348
 FT CARBOHYD 360 360
 FT CARBOHYD 405 405
 SQ SEQUENCE 1076 AA; 123119 MW; 9B53F16E05E80EB9 CRC64;

Query Match 9.4%; Score 106; DB 1; Length 1076;
 Best Local Similarity 19.6%; Pred. No. 2;
 Matches 47; Conservative 37; Mismatches 78; Indels 78; Gaps 10;

24 KSTNTVHKSSEAGTSVFYKKTGDMLEPDTTHVWFLINNEKSYVSKDITIKQIOGG 82
 179 KASNLEPKFSFNTSYVF--KNG-----TSEHCFTVIMNLEAGVSYFSQVLEKEMLRON 232
 83 QQLD-----LSTLININVTGTH-----SNVYSGQSAITDF 111
 233 EELQKILKDPNRRSNIVMCGIPQTMESLKIDWTATEDYVILLVDLFNNVYLEENATAP- 291
 112 EKAFPSKITTVDNKTITIVTTPQGGTSGNSINIKTKITNEQCKEFPVNSQAW----- 166
 292 -----DYMKLVLLTLPNGSTIN-----TSLSKESLQEFDFALAYLDGL 333
 167 -----YOEHGKEEVNSKSNHYHNINANAGI-----EGYKGLKYLKODKDK 211
 334 LGHMLKTLRLNG-ENTTAHKAFAHARNLTFEGSGTGPVTLDDSGDIDNRMVLLVTSVDK 392

RESULT 3
 ATL_STAUV STANDARD; PRT: 1256 AA.
 ID ATL_STAUV
 AC P52081;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE BIOFUNCTIONAL AUTOLYSIN PRECURSOR (INCLUDES: N-ACETYLORAMOYL-L-ALANINE
 DE AMIDASE (EC 3.5.1.28); MANNOSYL-GLYCOPROTEIN ENDO-BETA-N-
 DE ACETYLGALUCOSAMIDASE (EC 3.2.1.96)).
 GN ATL.

OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
 RC STRAIN-EN450;
 RX MEDLINE=95116542; PubMed=7816834;
 RA Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suglnaka H.,
 RA Tomasz A.;
 RT "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
 alanine amidase domain and an endo-beta-N-acetylglucosaminidase
 domain: cloning, sequence analysis, and characterization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MCTC 8325-4;
 RA Foster S.J.;
 RT Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOSOYL UNIT IN
 HIGH-MANNOSE GLYCOPROTEIDS AND GLYCOPROTEINS CONTAINING THE
 -[MAN)5(GLCNAc)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
 RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
 OLIGOSACCHARIDE IS RELEASED INTACT.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES THE LINK BETWEEN N-ACETYLMURAMOYL
 RESIDUES AND L-AMINO ACID RESIDUES IN CERTAIN BACTERIAL CELL-WALL
 GLYCOPROTEIDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
 EXTRACELLULAR LYTIC ENZYMES.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
 ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
 GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D17366; BAA04185.1; -
 DR EMBL: L41499; AAA99982.1; -
 DR InterPro: IPR002502; Amidase_2.
 DR InterPro: IPR002901; Amidase_4.
 DR Pfam: PF01510; Amidase_2; 1.
 DR Pfam: PF01832; Amidase_4; 1.
 DR SMART: SM00047; LY22; 1.
 DR Cell wall; Hydrolyase; Signal; Multifunctional enzyme; Repeat.
 KW SIGNAL
 FT CHAIN 1 29
 FT SIGNAL 1 29
 FT CHAIN 30 1256
 FT SIGNAL 30 1256
 FT DOMAIN 139 775
 FT DOMAIN 776 1256
 FT DOMAIN 425 589
 FT REPEAT 536 758
 FT REPEAT 770 932
 FT REPEAT 1256 137384
 FT SEQUENCE 1256 AA; 137384 MW; 2BB76CA292FDD20 CRC64;
 Query Match 8.7%; Score 98; DB 1; Length 1256;
 Best Local Similarity 21.6%; Pred. No. 9.5;
 Matches 61; Conservative 39; Mismatches 90; Indels 92; Gaps 16;
 OY 15 KYATISGKSTVTVYHKEAGTSVFE---YKTKGD-----MLPEPTTH---VRWF 59
 DB 790 KTAQVKNPTGIRASYETKAKNGAKYADRTFYTKERAHGNTTYLLNNTSINILGWF 849
 OY 60 ---ININN-EKSYVSKDITKDOIQGOQDLSTLN-INVTGHSNYSQGSATDFE 112
 DB 850 NVADLNVQNLGKEVKTQYKTYVKNKSNNGLSMVPWGTKNQVILLGN---IAQGTNARK 905

OY 113 KAPGSKI---TVDNWK---NTIDVTIP-----OGYGSY-----N 141
 DB 906 QSVGVGADVLYIGTIRNNRTGTVNNAKDLAPYAVKPTISAADKYVTVYKNGNYYVTPN 965
 OY 142 SFESINRYKTKITNEQO---KEFYVNSQAWYQEHGKEEVNCK-----SENNIT 183
 DB 966 SDTAKYSLKAFNEQPPAVVEQVINGQTWY--YGKLS-NGKLAMIKSTDLAKELIKYNOF 1022
 OY 184 VHNINANAGIEGTGVKELKY-----LQDKRTK 211
 DB 1023 GMTLNOVAQIQAGLYQKPPQVRVGRKWTDAKFNVDVXHAMDTK 1064
 RESULT 4
 ID TOP1_BORBU STANDARD; PRT; 848 AA.
 AC 051768;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)
 DE (UNWINDING ENZYME) (SWIHELASE).
 GN TOPA OR BB0828.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathgr R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwin M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utechtack T., Matthey L., McDonald L., Artlich P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi.
 RL Nature 390:580-586(1997).
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
 CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
 CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
 DNA, FOLLOWED BY PASSAGE AND REJOINING.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
 BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
 WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
 AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE001180; AAC67161.1; -
 DR TIGR: BB0828; -
 DR InterPro: IPR003601; DNATopI_ATP_bind.
 DR InterPro: IPR003602; DNATopI_DNA_bind.
 DR InterPro: IPR000380; Pro_topoisomase.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF01131; Topoisom_Dac; 1.
 DR Pfam: PF01751; Toprim; 1.
 DR PRINTS: PR00417; PRPISMRASEI.
 DR SMART: SM00437; TOP1AC; 1.
 DR SMART: SM00436; TOP1BC; 1.
 DR SMART: SM00493; TOP1RM; 1.

RC STRAIN-ATCC 9790;
 RX MEDLINE-92165737; PubMed-1347040;
 RA Chu C.-P., Kariyama R., Daneo-Moore L., Shockman G.D.;
 RT "Cloning and sequence analysis of the muramidase-2 gene from
 RT Enterococcus hirae.";
 RL J. Bacteriol. 174:1619-1625(1992).
 RN (2)
 RP FUNCTION.
 RC STRAIN-ATCC 9790;
 RX MEDLINE-89327152; PubMed-2753858;
 RA Dolinger D.L., Daneo-Moore L., Shockman G.D.;
 RT "The second peptidoglycan hydrolase of Streptococcus faecium ATCC 9790
 RT covalently binds penicillin.";
 RL J. Bacteriol. 171:4355-4361(1989).
 RN (3)
 RP FUNCTION.
 RC STRAIN-ATCC 9790;
 RX Del Mar Lleo M., Canepari P., Satta G.;
 RT "Thermosensitive cell growth mutants of Enterococcus hirae that
 RT elongate at non-permissive temperature are stimulated to divide by
 RT parental autolytic enzymes.";
 RL J. Gen. Microbiol. 139:3099-3117(1993).
 CC -1- FUNCTION: MAY WORK IN CONCERT WITH AND POTENTIATE THE PROGRESSIVE
 CC HYDROLYTIC ACTION OF MURAMIDASE-1, WHICH REQUIRES BINDING OF THE
 CC ENZYME TO NONREDUCING ENDS OF GLYCAN CHAINS. HYDROLYSIS IN THE
 CC MIDST OF GLYCAN CHAINS WOULD INCREASE THE NUMBER OF BINDING SITES
 CC FOR MURAMIDASE-1. MAY FUNCTION IN FACILITATING SEPTUM FORMATION
 CC AND CELL SEPARATION. ACTIVE ON M. LUTEUS CELL WALLS AND ON E. HIRAE
 CC CELL WALL FRACTIONS, BUT NOT ACTIVE ON E. HIRAE IMPACT CELL WALLS.
 CC CAN COVALENTLY BIND PENICILLIN.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN N-
 CC ACETYL-D-GLUCOSAMINE AND N-ACETYL-MURAMIC ACID IN PEPTIDOGLYCAN
 CC HETEROPOLYMERS OF THE PROKARYOTES CELL WALLS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC BINDING.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 6 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M77639; AAA24776.1; -
 CC DR PIR: A42296; A42296.
 CC DR HSSP: P21697; 1PCS.
 CC DR InterPro: IPR002901; Amidase_4.
 CC DR InterPro: IPR002482; LysM.
 CC DR Pfam: PF01832; Amidase_4; 1.
 CC DR Pfam: PF01476; LysM; 6.
 CC DR SMART: SM00257; LysM; 6.
 CC DR SMART: SM00047; LysM; 1.
 CC KM Hydrolyase; Glycosidase; Bacteriolytic enzyme; Cell wall;
 CC Cell division; Septation; Repeat; Signal.
 CC KW SIGNAL
 CC FT CHAIN 1 49
 CC FT REPEAT 50 666 MURAMIDASE-2.
 CC FT REPEAT 257 299 LYSM 1.
 CC FT REPEAT 338 380 LYSM 2.
 CC FT REPEAT 414 456 LYSM 3.
 CC FT REPEAT 489 531 LYSM 4.
 CC FT REPEAT 565 607 LYSM 5.
 CC FT REPEAT 623 665 LYSM 6.
 CC SEQUENCE 666 AA; 70670 MW; FFOA7FAFCD810BA3 CRC64;

Query Match 8.5%; Score 95; DB 1; Length 666;
 Best Local Similarity 24.0%; Pred. No. 7.2;
 Matches 49; Conservative 28; Mismatches 81; Indels 46; Gaps 11;

OY 20 TSGNKTNTVTHKSEAGTSVFYKTDGMLPEDTTH-----VRMFLINNEKSYVS 70
 DB 242 TSGNKGSGSATL-----TGTTYVKSQDSV-WGISHSFGITMAQLLEW-NNKNKNTYFP 293
 OY 71 KDITIKDQIOGGQOLDLSTLNI-----NVYG-THSNYSQGSATIDFEKAPPGSKITY 122
 DB 294 QKLLIK-----CGQSAGSSTNTGNNASSGNTSGNTNTSGTGAT-----GAKYTV 340
 OY 123 DNTKNTIDVTIPQGYGSYNSFSINFKTKITNEOQKEVNNNSQAMYDHEKVEVNGKSFNH 182
 DB 341 KSDGSVWKIKNDHC-----ISMNMLE-WNNIKNNVYPGQDLVYKSGSSASGTSNT 393
 OY 183 TVNINANGIEG-TYKGLKVLK 205
 DB 394 STGNTSNTANTSTTSGSTYTVK 417
 RESULT 7
 YDBA_ECOLI STANDARD; PRT; 2003 AA.
 ID YDBA_ECOLI
 AC P33666; P76087; P76088; P76856; P76857; P76859;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 GN YDBA OR B1401/B1405.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-Kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 464-2003 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-92190338; PubMed-1665986;
 RA Moszer I., Glaser P., Danchin A.;
 RT "Multiple insertion sequences near the replication terminus in
 RT Escherichia coli K-12.";
 RL Biochimie 73:1361-1374(1991).
 CC -1- SIMILARITY: TO S.TYRPHIMORIUM ORF NEAR CYSG (AC P25928).
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
 CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
 CC BETWEEN AMINO ACIDS 839 AND 840.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----
 DR EMBL: AE000237; AAC74483.1; ALT_SEQ.
 DR EMBL: AE000237; AAC74487.1; ALT_SEQ.
 DR EMBL: D90778; BAA15009.1; ALT_SEQ.
 DR EMBL: D90778; BAA18880.1; ALT_SEQ.
 DR EMBL: D90779; BAA18881.1; ALT_SEQ.
 DR EMBL: X62680; NOT_ANNOTATED_CDS.
 DR Ecogene: B61307; ydda.
 DR Hypothetical protein; Complete proteome.
 KM CONFLICT 489 489 I -> V (IN REF. 2).
 FT CONFLICT 495 495 I -> V (IN REF. 2).
 SQ SEQUENCE 2003 AA; 205949 MW; B83A12C6B53220BE CRC64;

Query Match 8.5%; Score 95; DB 1; Length 2003;
 Best Local Similarity 25.1%; Pred. No. 28;
 Matches 69; Conservative 30; Mismatches 92; Indels 84; Gaps 16;
 QY 5 HHHHHG-----SDDKATITSGNK-----STNYVYAKSEA 35
 DB 94 HHHRRNSPLRPPTPPDESDPTVPPTGGDEIIPDDPDPPTPPPKFVSFNNVYIDKTEK 153
 QY 36 GNS-----SVFY---KTGDLPEDTTHVRFELNI-----NNEKSY--VSKDITIKDIOG 81
 DB 154 TITFINDSVFTYENADGTISLSDSNCRKATINLMQIDEANNVALEGVASADGATKQVNH 213
 QY 82 GOOL-----DLSTLIN-----VTGTHSNVYSGQSAITPEK-----AFPGSK 119
 DB 214 NCELVTGTGNATVNNNGKTVGDKDSTGEINGNCK-VIOGGDLVSGGHCIDITDS 272
 QY 120 ITVDNKTNTIDVTIQGYC---SYNSFSINT--KTKITNEQKEFYNNQAMQOEKKE 174
 DB 273 ATVDN-KGTMTVTDPESMGIDGDKAIVNNEGSEFTTGTGTQINGDADATANNNGKTT 331
 QY 175 VNGKSEHTVHNINANAGIEGTVGKELVKYKODK 209
 DB 332 VDGKDSGT--EINGNG-----KVI-ODGD 354

RESULT 8
 BXG_CLOBO STANDARD; PRT: 1296 AA.
 AC 060193;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BOTULINUM NEUROTOXIN TYPE G PRECURSOR (EC 3.4.24.69) (BONT/G)
 GN (BONTOKILYSIN G).
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=113 / 30;
 RX MEDLINE=94092745; PubMed=8268233;
 RA Campbell K., Collins M.D., East A.K.;
 RT "Nucleotide sequence of the gene coding for Clostridium botulinum
 (Clostridium argentinense) type G neurotoxin: genealogical comparison
 with other clostridial neurotoxins.";
 RL Biochim. Biophys. Acta 1216:487-491(1993).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 ENDOPEPTIDASE.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
 WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
 FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).

CC -----
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
 BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 SUPERFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb.ch).
 CC
 DR EMBL: X74162; CAA52275.1; -.
 DR HSSP: P04958; IAF9.
 DR MEROPS: M27.002; -.
 DR InterPro: IPR000395; Bontokilysin.
 DR InterPro: IPR000130; Zn_Mrppeptase.
 DR Pfam: PF01742; Peptidase_M27; 1.
 DR PRINTS: PR00760; BONTOKILYSIN.
 DR Prodom: PD001963; Bontokilysin; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KM Neurotoxin; Hydrolase; Metalloprotease; Zinc.
 FT INIT_MET 0
 FT CHAIN 1 441
 FT METAL 442 1296
 FT ACT_SITE 229 229
 FT ACT_SITE 230 230
 FT METAL 233 233
 FT DISULFID 435 449
 SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F66531 CRC64;

Query Match 8.3%; Score 93.5; DB 1; Length 1296;
 Best Local Similarity 23.0%; Pred. No. 21;
 Matches 50; Conservative 32; Mismatches 66; Indels 69; Gaps 12;
 QY 47 DMLP-----EDTTHVRFELNIINNEKSYVSKDITIKDIOGGQOOLDS-----TLNINVT 95
 DB 838 DSIPLPDLISLYTKDTILIOVF---NNYISMISSNAILSLSYRGRRLDSSGYGATMNVGSD 894
 QY 96 GTHSNVYSGQSAITPEKAPGSKITVDNKTNTIDVTIQGYGS-YNSFSINT----- 147
 DB 895 VLFNDIGNQOFKLINSE-----NSNITAHQSKFVY-----YDSMFDNFSINFWATPKY 943
 QY 148 -KTKITNEQKE-----VNNQSAW-YQEHKE-----EVNGKS----- 179
 DB 944 NNNDIOFTYLDNEVYTIISCKRDSGKMSVSIKGNRIITWTLIDVNAKSKSIPEYSIKDNISD 1003
 QY 180 -----FNHTVHNIN--NANAGIEGTVGKELVKYKODK 208
 DB 1004 YINKWFSTITITDRLGANANIYINGSLAKSEKILINDR 1040
 RESULT 9
 TOP1_MYCPN STANDARD; PRT: 711 AA.
 ID TOP1_MYCPN
 AC P78032;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)
 GN (UNWISTING ENZYME) (SWIVELASE).
 GN TOPA OR MPR261 OR MP572.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=9710585; PubMed=8948633;

Db 131 FRPYIGARVAVGHVHRSIDSTKKTIEVTYPSNAPNGAVTTNTDPKT-----QNDYOS 184

QY 162 NS 163

Db 185 NS 186

RESULT 11

GUN_BACSP STANDARD; PRT; 463 AA.

ID GUN_BACSP

AC P29019;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE) (ENDO-K).

OS Bacillus sp. (strain KSM-330).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxId=1409;

RP SEQUENCE FROM N.A.

RA MEDLINE=92121880; PubMed=1770347;

RT Ozaki K., Sumitomo N., Ito S.;

RT "Molecular cloning and nucleotide sequence of the gene encoding an endo-1,4-beta-glucanase from Bacillus sp. KSM-330.";

RL J. Gen. Microbiol. 137:2299-2305(1991).

RN [2]

RP SEQUENCE OF 56-75, AND CHARACTERIZATION.

RX MEDLINE=91259037; PubMed=2045781;

RA Ozaki K., Ito S.;

RT "Purification and properties of an acid endo-1,4-beta-glucanase from Bacillus sp. KSM-330.";

RL J. Gen. Microbiol. 137:41-48(1991).

CC -1- FUNCTION: THIS ACID ENDOGLUCANASE IS ACTIVE OVER AN EXTREMELY NARROW RANGE OF PH VALUES, BETWEEN 4.5 AND 6.5, WITH AN OPTIMUM PH AT 5.2.

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- PTM: THE N- AND THE C-TERMINUS MAY BE SUBJECTED TO PROTEOLYSIS.

CC -1- MISCELLANEOUS: ONE TRP RESIDUE HAS BEEN PROVED TO BE INVOLVED IN THE MECHANISM OF ACTION OF ENDO-K.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL: M68872; AAA22409.1; -.

CC PIR: A44808; A44808.

DR HSSP: P04955; ICEM.

DR InterPro: IPR002037; Glyco_hydro.8.

DR Pfam: PF01270; Glyco_hydro.8; 1.

DR PRINTS: PR00735; GLHYDRLASE8.

DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.

CC Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 ?

FT PROPEP 55 ?

CC POTENTIAL.

CC ENDUGLUCANASE.

FT CHAIN 56 463

FT ACT_SITE 130 130

FT ACT_SITE 191 191

CC NUCLEOPHILE (POTENTIAL).

CC SEQUENCE 463 AA; 51882 MW; 407FA54F5236C59E CRC64;

SO

Query Match 8.1%; Score 91; DB 1; Length 463;

Best Local Similarity 21.2%; Pred. No. 9;

Matches 53; Conservative 31; Mismatches 64; Indels 102; Gaps 13;

QY 7 HHHHSGDDKY-----ATTSGNKSTNTV-----VHSEAGTSSVFEYKGTDMIPED--T 53

Db 204 HKQWSSGSKINTYKEAQNMTTKGIRKASVTKRNGNLGMDGSTF-----DTRPSDMM 238

QY 54 THYWFNLINNEKSYSGKDIITKQIGGOGLDSTLNI--NTYGTSHNY---YSGOSA- 107

Db 259 SHIRAFYEFTGDKT-----LNVIDLKXNTYTNFTNKSPTKGL 297

QY 108 ITDF---EKAPEGSKITVDNTKNT-----IDVTIPQYGSYN-----SFS 144

Db 298 ISDFVYKNNPPQAPKDFLDSESKYTDSYYNARSVPKLRIVMDYAMYGKRGKISDKVATM 357

QY 145 INKTK-----ITNEQKEFVNNNSAHWQEG 171

Db 358 IKSITKCNPSKIVDGYKLDSTNIGDYPTAVYSPFIAGTTSKNDQWVNSGMDW-MKNK 416

QY 172 KEENYNGKSFN 181

Db 417 KESYFSDSYN 426

RESULT 12

LIPAB_MYCPU STANDARD; PRT; 578 AA.

ID LIPAB_MYCPU

AC Q50274;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE LIPOPROTEIN A PRECURSOR.

GN LIPAB OR MYPU_5300.

OS Mycoplasma pulmonis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxId=2107;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KD735-15;

RX MEDLINE=20245550; PubMed=10781561;

RA Shen X., Gumulak J., Yu H., French C.T., Zou N., Dybvig K.;

RT "Gene rearrangements in the *usa* locus of *Mycoplasma pulmonis*.";

RL J. Bacteriol. 182:2900-2908(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=QAB CTIP;

RX MEDLINE=21267165; PubMed=11353084;

RA Chandaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F., Moszer I., Dybvig K., Wroblewski H., Viarl A., Rocha E.P.C., Blanchard A.;

RT "The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*.";

RL Nucleic Acids Res. 29:2145-2153(2001).

RN [3]

RP PARTIAL SEQUENCE FROM N.A.

RC STRAIN=KD735-15;

RX MEDLINE=96414471; PubMed=8817492;

RA Bhugra B., Voelker L.L., Zou N., Yu H., Dybvig K.;

RT "Mechanism of antigenic variation in *Mycoplasma pulmonis*: interwoven, site-specific DNA inversions.";

RL Mol. Microbiol. 18:703-714(1995).

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE PULMONIS LIPAB LIPOPROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL: U23947; AABA1030.2; -.

DR EMBL: AL445565; CAC13703.1; -.

DR mypulst: MYPU_5300: -
 KW Lipoprotein; Membrane; Signal; Complete proteome.
 FT SIGNAL 1
 FT CHAIN 27
 FT LIPID 28 578
 FT DOMAIN 82 165
 FT VARIANT 7 7
 FT VARIANT 41 41
 FT VARIANT 110 110
 FT VARIANT 127 127
 FT VARIANT 547 547
 SQ SEQUENCE 578 AA; 66219 MW; B533ED3467005D89 CRC64;

Query Match 8.1%; Score 91; DB 1; Length 578;
 Best Local Similarity 20.2%; Pred. No. 12;
 Matches 49; Conservative 33; Mismatches 103; Indels 58; Gaps 6;

QY 14 DKVATITSGN-STNVYHKESEAGTSVFYKGTGDMLEPETHVRWFLNINNEKSYSKD 72
 DB 61 DKXNVSDNDKSTNKA VNSNENOTOKNESONTKDDSKTSNLTTONSSNTKSK 120
 QY 73 ITIKDIQGGQDLSTLINVGTSHSNYSGSAITDFEKAPEGSKITVDNTRNTIDVT 132
 DB 121 IQENKQ---SQKQNTSAVNSALE-----KQTKNDENISLVNSKDTNYILKNDKVA 170
 QY 133 IPQGTG---SYNSFSINKTKITNEQKEFYVNSQ---AWYOE----- 169
 DB 171 LAKDSEKSKSNLSNLTNKPVENRONKNEVADKALQWQKINESASILESPSYDQTS 230
 QY 170 -----HGKEVNGKSFNHHVHNINNAGIEGTVEKELKX 205
 DB 231 LSLTFKEGMPLEVLVLEKLEMDSHHEKEISFKITNGKVNVLLTSSNLISGKMKINSFS 290
 QY 206 QDK 208
 DB 291 FDK 293

RESULT 13
 C1GB_BACT2 STANDARD; PRT; 1169 AA.
 AC 09ZAZ6;
 DT 30-MAY-2000 (Rel. 39, Last Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PESTICIDIAL CRYSTAL PROTEIN CRYGB (INSECTICIDIAL DELTA-ENDOTOXIN
 GN CRYGB) (CRYSTALINE ENTOMOCIDAL PROTOXIN) (133 KDA CRYSTAL PROTEIN).
 OS Bacillus thuringiensis (subsp. wuhanensis).
 OC Bacillus; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 OX NCBI_Taxid=52024;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HD-525;
 RX MEDLINE-20133386; PubMed-10688690;
 RA Kao W.-S., Lin J.-H., Tzeng C.-C., Kao S.-S., Chak K.-F.;
 RT "Cloning of two new cry genes from Bacillus thuringiensis subsp.
 RT wuhanensis strain";
 RT Curr. Microbiol. 40:227-232(2000).
 CC -1- FUNCTION: PROMOTES COLLOIDOSMOTIC LYSIS BY BINDING TO THE MIDGUT
 CC EPITHELIAL CELLS OF LEPIDOPTERAN LARVAE. TOXIC TO PIERIS RAPAE.
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
 CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
 CC OF THE SPORE COAT.
 CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
 CC N-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U70725; AAD10291.1; -
 DR HSSP: P02965; ICIT.
 DR InterPro: IPR001178; Endotoxin.
 DR Pfam: PF00555; endotoxin; 1.
 KW Toxin; Sporulation.
 SQ SEQUENCE 1169 AA; 132904 MW; D1EFC1508A8B10BD CRC64;

Query Match 8.1%; Score 90.5; DB 1; Length 1169;
 Best Local Similarity 24.5%; Pred. No. 31;
 Matches 36; Conservative 17; Mismatches 81; Indels 13; Gaps 5;

QY 28 VYHKESEAGTSVFYKGTGDMLEPETHVRWFLNINNEKSYSKDITIKDIQGGQDL 87
 DB 1021 VYAKREGGECVTHET-----EDNDELKFSNCEVEEITPNNTVTCNDTATQEE- 1074
 QY 88 STLINVTGTHSNYSGSAITDFEKAPEGSKITVDNTRNTIDVTIPQYGSYNSFSINY 147
 DB 1075 GTYTSRNNGYDGAESNSVPADYASAEKAYTPDGRDNTCESN--RGYGDYTPLPAGY 1132
 QY 148 KTKITNEQKEFYVNSQAWYOEHKKE 174
 DB 1133 VTK---ELEVEPETHKWI-EIGETE 1154

RESULT 14
 ETL1_CAEL STANDARD; PRT; 416 AA.
 ID ETL1_CAEL
 AC P28515;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSCRIPTION FACTOR ETL-1.
 GN ETL-1 OR W09C2.1.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91342668; PubMed-1875944;
 RA Spieth J., Shim Y.H., Lea K., Conrad R., Bjumenthal T.;
 RT "etl-1, an embryonically expressed Caenorhabditis elegans gene
 RT homologous to the GATA transcription factor family.";
 RT Mol. Cell. Biol. 11:4651-4659(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Lloyd C.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF GENES CONTAINING THE GATA
 CC REGION, INCLUDING VITELLOGENIN GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X57834; CAA40967.1; -
 DR EMBL: Z68221; CAA92494.1; ALT_INT.
 DR PIR: A41267; A41267.
 DR HSSP: P17678; IGAU.
 DR WormPep: W09C2.1; CE03799.

DR InterPro: IPR000679; ZnF_GATA.
 DR Pfam: PF00320; GATA_2.
 DR PRINTS: PR00619; GATA2NFINGER.
 DR SMART: SM00401; ZnF_GATA_2.
 DR PROSITE: PS00344; GATA_ZN_FINGER_1; 2.
 DR PROSITE: PS00114; GATA_ZN_FINGER_2; 2.
 DR Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN_FING 217 241 GATA-type 1.
 FT ZN_FING 272 296 GATA-type 2.
 SQ SEQUENCE 416 AA; 44823 MW; 01FF5A19A57D42D CRC64;

Query Match 8.0%; Score 89.5; DB 1; Length 416;
 Best Local Similarity 22.0%; Pred. No. 10;
 Matches 45; Conservative 34; Mismatches 89; Indels 37; Gaps 9;

QY 11 GSDGVATITSGKSNVYHKSSEAGTSVFYKGTGDMLEPDTTHRMFLINNKSYVS 70
 DB 13 GISSGGASLAPSTTASLAPSYNTSATVNT-----TPSSYPMTFNYO---YA 61
 QY 71 KDITKDIQGGQQLDLSTLINVTGTHSN--YSGQSAITDF-----EKAPFGSKITVD 123
 DB 62 GGTVTYTDMAFGSGMDMS-MNNGVFGQNNPSYFPTTOLNTYGYDTLAATAAGITVN 120
 QY 124 NIKNTIDVTIPQGYST--NSFSINKTKITNQKEEVNNSQ-----AMYQEHKE 173
 DB 121 N-NOINVTIVQNGTIVPQITQNIITSVSNVQSSVPIINNSQPLMPTGLACGTS 178
 QY 174 EVNGKSPN-----HTVHNINANAG 192
 DB 179 SASSSSANSTSPKNTISKANRSSG 203

RESULT 15
 LAC2_THACU STANDARD: PRT; 599 AA.
 AC 002075;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LACCASE 2 PRECURSOR (EC 1.10.3.2) (BENZENEDIOL:OXYGEN OXIDOREDUCTASE)
 DE (URSHIOL OXIDASE) (DIPHENOL OXIDASE).
 GN LCC2.
 OS Thanaeophorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;
 OC mitosporic Ceratobasidiaceae; Rhizoctonia.
 OX NCBI_TaxID=107832;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R22;
 RC MEDLINE=96171523; PubMed=8598061;
 RA Walethner J.A., Xu F., Brown K.M., Brown S.H., Gollightly E.J.,
 RA Halkier T., Kauppinen S., Pederson A., Schneider P.;
 RT "The identification and characterization of four laccases from the
 RT plant pathogenic fungus Rhizoctonia solani.";
 RT Plant. Genet. 29:395-403(1996).
 RL Curr. Genet. 29:395-403(1996).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE +
 CC 2 H(2)O.
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: IN MICELIA, AT A LOWER LEVEL THAN LCC4.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z54276; JAS.1041.1; -.
 DR HSSP: P37064; IASP.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; MultiCu-oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
 DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
 KW oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
 KW Lignin degradation; Multigene family.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 599 LACCASE 2.
 FT DOMAIN 21 145 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 157 307 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 450 567 PLASTOCYANIN-LIKE 3.
 FT METAL 82 82 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 84 84 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 127 127 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 497 497 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 500 500 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 502 502 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 549 549 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 550 550 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 551 551 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 551 551 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 551 551 COPPER (TYPE 1) (BY SIMILARITY).
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 599 AA; 66540 MW; F01DD5C52D0CB91 CRC64;

Query Match 7.9%; Score 89; DB 1; Length 599;
 Best Local Similarity 22.7%; Pred. No. 17;
 Matches 60; Conservative 28; Mismatches 104; Indels 72; Gaps 15;

QY 3 GSHHHHGGSDKVAATITSGNKSNTYVHKSSEAGTSVFYKGTGDMLEPDT--THRMFL 60
 DB 122 GTYWHSHLSQYV---DLRGRPIVYDHPDHYRN---YDVDD--ERTVFTLADWY- 170
 QY 61 NINNKSYVSKDI--TKQD--IQGGQQLDLSTLINVTGTHSNYS----- 103
 DB 171 HTPESEALIAITHDVAKTIPDSGTINGKGYKDPASANTNNT-TLENLYLKYKRRKRLRI 229
 QY 104 -GQSATIDFEKAPFGSKITVDN-----TK-----NTIDVTIPQGYG-----SY--- 140
 DB 230 INASALASFFGVQGHKCTIIEADVLTIRPIEVDADILAGQVTSCLKADQDPDSWIN 289
 QY 141 ---NSFSINKTKITNQKE-----FYVNS--QAMYQEHKEEVNGSSEFN 182
 DB 290 APTTNVLTNTVQALLVYEDDKRPTHYFWKPFILWTKISNETIQQWQHKGSHGKGGHHH 349
 QY 183 TVNINANAGIECTVAGELVKLO 206
 DB 350 KVALIGVSGLSRVKVRASDLK 373

Search completed: January 30, 2002, 00:11:31
 Job time: 544 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 00:01:12 ; Search time 91.71 Seconds
(Without alignments)
336.533 Million cell updates/sec

Title: US-09-813-820-4
Perfect score: 1122
Sequence: 1 MNGSHHHHNSGDKVAVTIT.....GIEGVKGLKYLKDKDKTK 211

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.potent:*
13: sp.virus:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	182	16.2	319	2	09F859	09f859 enterococcus
2	181	16.1	721	2	09F866	09f866 enterococcus
3	179	16.0	627	2	09F863	09f863 enterococcus
4	178	16.0	721	2	09F867	09f867 enterococcus
5	178	15.9	319	2	09F857	09f857 enterococcus
6	178	15.9	319	2	09F856	09f856 enterococcus
7	178	15.9	458	2	09F870	09f870 enterococcus
8	178	15.9	580	2	09F862	09f862 enterococcus
9	178	15.9	580	2	09F861	09f861 enterococcus
10	178	15.9	627	2	09F864	09f864 enterococcus
11	178	15.9	674	2	09F865	09f865 enterococcus
12	178	15.8	319	2	09F858	09f858 enterococcus
13	175	15.6	296	2	09F860	09f860 enterococcus
14	175	15.6	319	2	09F860	09f860 enterococcus
15	175	15.6	1092	2	09F860	09f860 enterococcus
16	175	15.6	1092	2	09F860	09f860 enterococcus
17	131	11.7	1092	2	09F860	09f860 enterococcus
18	105.5	9.4	598	2	09JPR7	09jpr7 lactococcus
19	100.5	9.0	595	2	09JPH0	09jph0 neisseria m

20	100.5	9.0	599	2	09JPS8	09jps8 neisseria m
21	100	8.9	1536	2	048031	048031 haemophilus
22	99.5	8.9	589	2	09JPI0	09jpi0 neisseria m
23	99.5	8.9	592	2	09JOM4	09jom4 neisseria m
24	99	8.8	1367	5	09V9M0	09v9m0 drosophila
25	99	8.8	1368	5	09NVJ6	09nvj6 drosophila
26	98.5	8.8	600	2	09JPS5	09jps5 neisseria m
27	98.5	8.8	4919	2	09ZHL0	09zhl0 haemophilus
28	98	8.7	949	5	09J306	09j306 plasmidium
29	98	8.7	1117	2	053971	053971 streptococc
30	98	8.7	1248	2	099V41	099v41 staphylococ
31	97	8.6	431	2	09RNM5	09rmn5 bacillus ce
32	97	8.6	1181	5	09J260	09j260 plasmidium
33	96.5	8.6	1733	2	09K114	09k114 staphylococ
34	96	8.6	392	6	028143	028143 bos taurus
35	96	8.6	481	2	032391	032391 staphylococ
36	95	8.5	332	2	068170	068170 lactococcus
37	95	8.5	359	2	09CHR9	09chr9 lactococcus
38	95	8.5	799	2	09A117	09a117 photorhabdu
39	94.5	8.4	188	2	099R42	099r42 staphylococ
40	94.5	8.4	2902	2	09ZME6	09zme6 helicobacte
41	94.5	8.4	3890	2	099U53	099u53 staphylococ
42	94	8.4	2380	5	096266	096266 plasmidium
43	94	8.4	2678	5	09NDS4	09nds4 dictyosteli
44	93.5	8.3	532	10	09T069	09t069 arabidopsis
45	93.5	8.3	1946	5	09J291	09j291 plasmidium

ALIGNMENTS

RESULT	1	ALIGNMENTS
09F859	PRELIMINARY;	PRT; 319 AA.
ID	09F859;	
DT	01-MAR-2001 (TREMBLrel. 16, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE	COLLAGEN ADHESIN PRECURSOR (FRAGMENT).	
GN	ACE.	
OS	Enterococcus faecalis (Streptococcus faecalis).	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;	
OC	Enterococcus.	
OX	NCBI_TaxID=1351;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BEIRUT/TX0645;	
RX	MEDLINE=20407335; PubMed=10948146;	
RA	Nallapareddy S.R., Singh K.V., Duh R.W., Weinstein G.M., Murray B.E.;	
RT	"Diversity of ace, a gene encoding a microbial surface component	
RT	recognizing adhesive matrix molecules, from different strains of	
RT	enterococcus faecalis and evidence for production of ace during human	
RT	infections."	
RL	Infect. Immun. 68:5210-5217(2000).	
DR	EMBL: AF260889; AAC23947.1; -.	
FT	NON_TER 1	
FT	NON_TER 319	
SO	SEQUENCE 319 AA; 35986 MW; D5F914E0856A6940 CRC64;	
Query Match	16.2%; Score 182; DB 2; Length 319;	
Best Local Similarity	28.8%; Pred. No. 3.3e-06;	
Matches	55; Conservative 32; Mismatches 88; Indels 16; Gaps 6;	
QY	18 TITGKSTNVTVKSEAG--TSVFYKTKGDMLEDTTHVRFMLNINNEKSYVSKDIT 74	
DB	112 TATATGRLTIGVNTFETGQLEKRDYFPFYKVGDLAGE-SNQVRFLNVLNLSKSDVEDIS 170	
QY	75 IKDQIGGQQLDLSTLNINVTGHSNYSQGSATIDFEKAPGSGITVDNKNITDVIIP 134	
DB	171 IADRGSGQQLNKSKFPEIDVNDKRETKY--ISLAEPDQGYGKIDFV--TDNDFNLRFY 225	
QY	135 QGYGSYNSFSINVTYKTI--NEQCKEPVNSQAWYQEHGKEGVNGKSFNHTVHNINAG 192	

```

Db      226 QOKARTSFVTRTSTTEGQHQAATFENSYDINTYOLNNDATNEKNTSOV-----KNVF 280
Oy      193 IEGTVKGEKLV 203
Db      281 VEGEASGNQNV 291

RESULT  2
Q9F866  PRELIMINARY; PRT; 721 AA.
ID      Q9F866
AC      Q9F866;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-END6/7X0045;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL; AF260873; AAG23932.1; -.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      Pfam; PF00746; Gram_pos_anchor; 1.
DR      PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 79047 MW; F09EDBE49E5DF866 CRC64;

Query Match      16.1%; Score 181; DB 2; Length 721;
Best Local Similarity 28.8%; Pred. No. 1e-05;
Matches 55; Conservative 32; Mismatches 88; Indels 16; Gaps 6;

Oy      18 TTSGNKSNTVYHKSEAG---TSSVFYKTKGMDLPEDTTHVRFNLINNEKSYVSKDIT 74
AC      Q9F867;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL; AF260872; AAG23931.1; -.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      Pfam; PF00746; Gram_pos_anchor; 1.
DR      PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 78842 MW; 1CA8AID98CE69392 CRC64;

Oy      193 IEGTVKGEKLV 203
Db      321 VEGEASGNQNV 331

RESULT  3
Q9F863  PRELIMINARY; PRT; 627 AA.
ID      Q9F863
AC      Q9F863;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.

```

```

OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-MD9/7X0249;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL; AF260877; AAG23935.1; -.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      Pfam; PF00746; Gram_pos_anchor; 1.
DR      PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 627 AA; 68848 MW; 78C5345DFD99F528 CRC64;

Query Match      16.0%; Score 179; DB 2; Length 627;
Best Local Similarity 28.3%; Pred. No. 1.2e-05;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

Oy      18 TTSGNKSNTVYHKSEAG---TSSVFYKTKGMDLPEDTTHVRFNLINNEKSYVSKDIT 74
AC      Q9F867;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL; AF260872; AAG23931.1; -.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      Pfam; PF00746; Gram_pos_anchor; 1.
DR      PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 78842 MW; 1CA8AID98CE69392 CRC64;

Oy      193 IEGTVKGEKLV 203
Db      321 VEGEASGNQNV 331

RESULT  4
Q9F867  PRELIMINARY; PRT; 721 AA.
ID      Q9F867
AC      Q9F867;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL; AF260872; AAG23931.1; -.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      Pfam; PF00746; Gram_pos_anchor; 1.
DR      PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 78842 MW; 1CA8AID98CE69392 CRC64;

```

Query Match	16.0%	Score 179;	DB 2;	Length 721;
Best Local Similarity	28.3%	Pred. No. 1.5e-05;		
Matches 54;	Conservative 33;	Mismatches 86;	Indels 16;	Gaps 6;
Qy 18	TITSGNKSTNTVHKSEAG---TSSVFYKTKGDMIPEDTTHRWFLINLNNEKSYSKDIT 74			
Db 152	TATATGRLTIGVTMTERTGQIERDYPFPFYKVGDLAGE-SNGVRWFLNVLNNSDVTEDIS 210			
Qy 75	IKDQIGGQQLDLSLNLNNVTGTHSNNTYSGSALITDFEKAFPGSKITYDNTKNTIDVTIP 134			
Db 211	IADRGSSQQLNKESFTDIYNDKDKTKY---ISLAEFPOOGYGKIDFV--TNDENLRREY 265			
Qy 135	OGGYSGSYNSINYTKKIT--NBOQKEFNNQAGWQOEKGKEVNGKSFNHTYHNLNANAG 192			
Db 266	RNKAFRTFFIYRTYTITTEAGOHQATFENSIDINYQLNNQDATNEKNYSQV-----KNVF 320			
Qy 193	IEGTVGELKV 203			
Db 321	VEGEASGNQNV 331			
RESULT 5				
Q9F857	PRELIMINARY;	PRT;	319 AA.	
AC Q9F857;				
DT 01-MAR-2001 (TREMBLREL. 16, Created)				
DT 01-MAR-2001 (TREMBLREL. 16, last sequence update)				
DT 01-MAR-2001 (TREMBLREL. 16, last annotation update)				
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).				
GN AC.				
OS Enterococcus faecalis (Streptococcus faecalis).				
OC Bacteria; Filimicutes; Bacillus/Clostridium group; Enterococcaceae;				
OC Enterococcus.				
OX NCBI_TaxID=1351;				
RA SEQUENCE FROM N.A.				
RP STRAIN-PARKER/TX2619;				
RC MEDLINE=20407335; PubMed=10948146;				
RA Nallapareddy S.R., Singh K.V., Dub R.W., Weinstein G.M., Murray B.E.;				
RT "Diversity of ace, a gene encoding a microbial surface component				
RT recognizing adhesive matrix molecules, from different strains of				
RT enterococcus faecalis and evidence for production of ace during human				
RT infections.";				
RL Infect. Immun. 68:5210-5217(2000).				
DR EMBL: AF260894; AAC23952.1; -.				
FT NON_TER 1				
FT 319 319				
QO SEQUENCE 319 AA; 36001 MW; 2054C2B1IEC5920 CRC64;				

Query Match	15.9%	Score 178;	DB 2;	Length 319;
Best Local Similarity	28.3%	Pred. No. 6.5e-06;		
Matches	54;	Conservative	33;	Mismatches 88; Indels 16; Gaps 6;
QY	18	TITSNKNSTNNTVHKSSEAG---TSSVFYKTKTMDLPEDTHVRWFLNINNEKSYSKDTT	74	
DB	112	TATAAQRLLTIEGVNTEFGQIERDYPFEFKYVDLAGE-SNOYRWFLNVLNLSDDYTEDIS	170	
QY	75	IKDQIOGGQQLDLSTLNINVTGTHSNYSGGQALTDFFKAFPGSKITVDNKNTIDVTIP	134	
DB	171	IADRGSGQQLNKESFFTDIVNDKEFKY---ISLAFEGQGGKIDFV--TNDNFRLRY	225	
QY	135	QGYGYSNFSINRYKRTIT--NEQCKEFPVNSQAMTQEHGKEEYNGKSFPHLYVHNINANG	192	
DB	226	RDKARFTSFIVAYTSTITTEAGQHQTFFNSYDINQLNMQDATNEKNTSQV-----KNVF	280	
QY	193	IEGTVKGELK	203	
DB	281	VEGEASGNQNV	291	
RESULT	6			
	Q9F856			

ID	Q9F856	PRELIMINARY;	PRT;	319 AA.
AC	Q9F856.			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	COLLAGEN ADHESIN PRECURSOR (FRAGMENT).			
GN	ACE.			
OS	Enterococcus faecalis (Streptococcus faecalis).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;			
OC	Enterococcus.			
OX	NCBI_TaxID=1351;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-B-343/7X2783;			
RC	MEDLINE=20407335; PubMed=10948146;			
RA	Nallareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;			
RT	"Diversity of ace, a gene encoding a microbial surface component			
RT	recognizing adhesive matrix molecules, from different strains of			
RT	enterococcus faecalis and evidence for production of ace during human			
RT	infections.";			
RL	Infect. Immun. 68:5210-5217(2000).			
DR	EMBL; AF260896; AAC23954.1; -.			
FT	NON_TER	1		
FT	NON_TER	1		
FT	NON_TER	319		
SO	SEQUENCE	319 AA;	36073 MW;	94562E9DFD88988A CRC64;
Query Match	15.9%;	Score 178;	DB 2;	Length 319;
Best Local Similarity	28.3%;	Pred. No. 6.5e-06;		
Matches	54;	Conservative	33;	Mismatches 86;
				Indels 16;
				Gaps 6

```

Db 11.2 TATATGQRLIEGVTNMGQIREDXPFEFKVGDLAG- -SNQVRMFLNVLNKSQDVEDIS 170
Qy 75 IKDQIQGGQQLDLSTLNLNVGTGHSNYISGQSALIDFEKAFPGSKITVDNTKNTIDVTIP 134
Db 171 IADRGSGGQLKESFETDVIANDKETKY- -ISLAEFEQGYGKIDFV- -TNDNDFMLREF 225
Qy 135 QCGYSYNSPSINYKTKIT- -NEOQKEFVNSQAWQOEHOKEKEVNGCKSFPHHTYHNINANG 192
Db 226 ROKARFTSTSVRYRTSTTEAGQHQAATFENSYDINQNLNODATNEKNTSQV- -KNVF 280
Qy 193 IEGTVKSGELKY 203
Db 281 VEGEASGNQNV 291

RESULT 7
Q9EU70 ID Q9EU70 PRELIMINARY; PRT; 319 AA.
AC Q9EU70:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VARIOUS STRAINS;
RC MEDLINE=20407335; PubMed=10948146;
RA Nallapareddy S.R., Singh K.V., Dub R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
RT recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL, AF260895; AAC23953.1; -.
DR EMBL, AF260880; AAC23938.1; -.

```

DR EMBL: AF260881; AAG23939.1; -
 DR EMBL: AF260882; AAG23940.1; -
 DR EMBL: AF260883; AAG23941.1; -
 DR EMBL: AF260888; AAG23946.1; -
 DR EMBL: AF260890; AAG23948.1; -
 DR EMBL: AF260893; AAG23951.1; -
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 319 AA; 36031 MW; 2PCF56B1IAC90D20 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 319;
 Best Local Similarity 28.3%; Pred. No. 6.5e-06;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSNTVYHKSEAG---TSSVFYKKTGMDLPEDTTHVFMFLINNKESYSVKDIT 74
 DB 112 TATATGRLTEGVTNETGQIENDYFFFKVGLAGE-SNOQWFMFLNVLNKSVDYEDIS 170
 OY 75 IKDQIOGGQQLDLSTLINVYTHSNYSQSAITDPEKAFPSKITYDNTKNTIDVTIP 134
 DB 171 IADROSGGQQLNKESFTFIVNDKETRY--ISLAEFEQGYGKIDFV--TDNDFMLRFY 225
 OY 135 OGGSYNSFSINKTKIT--NEQOKFEVNNNSQAWYOEHEKKEEYNGKSFNHTVHNANAG 192
 DB 226 ROKARFTSFIVRTSTITTAGOHQATFENSYDINTYOLNODATNEKNTSOV-----KNVF 280
 OY 193 IEGTVKGLKV 203
 DB 281 VEGEASGNQNV 291

RESULT 8
 OYXB07 PRELIMINARY; PRT; 458 AA.
 AC OYXB07:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE COLLAGEN ADHESIN ACE (FRAGMENT).
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CG110;
 RA Rich R.L., Kreikemeier B., Owens R.T., Labrenz S., Narayana S.V.L.,
 RA Weinstock G.W., Murray B.E., Hook M.,
 RA "Ace: a collagen-binding MSCRAMM from Enterococcus faecalis."
 RL J. Biol. Chem. 0:0-0(1999).
 DR EMBL: AF159247; AAD43342.1; -
 DR HSSP: O53654; IAMX.
 DR COLLAGEN.
 KM
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 458 AA; 51085 MW; B674692BD29A0A24 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 458;
 Best Local Similarity 28.3%; Pred. No. 1e-05;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSNTVYHKSEAG---TSSVFYKKTGMDLPEDTTHVFMFLINNKESYSVKDIT 74
 DB 121 TATATGRLTEGVTNETGQIENDYFFFKVGLAGE-SNOQWFMFLNVLNKSVDYEDIS 179
 OY 75 IKDQIOGGQQLDLSTLINVYTHSNYSQSAITDPEKAFPSKITYDNTKNTIDVTIP 134
 DB 180 IADROSGGQQLNKESFTFIVNDKETRY--ISLAEFEQGYGKIDFV--TDNDFMLRFY 234
 OY 135 OGGSYNSFSINKTKIT--NEQOKFEVNNNSQAWYOEHEKKEEYNGKSFNHTVHNANAG 192

DB 235 ROKARFTSFIVRTSTITTAGOHQATFENSYDINTYOLNODATNEKNTSOV-----KNVF 289
 OY 193 IEGTVKGLKV 203
 DB 290 VEGEASGNQNV 300

RESULT 9
 OYF862 PRELIMINARY; PRT; 580 AA.
 AC OYF862:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LBJ-1/TX0020;
 RX MEDLINE=20407335; PubMed=10948146;
 RX Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.W., Murray B.E.,
 RT "Diversity of ace, a gene encoding a microbial surface component
 RT recognizing adhesive matrix molecules, from different strains of
 RT enterococcus faecalis and evidence for production of ace during human
 RT infections.";
 RL Infect. Immun. 68:5210-5217(2000).
 RL EMBL: AF260878; AAG23936.1; -
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor.1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Signal.
 FT SIGNAL. 1
 FT SIGNAL. 31
 FT SIGNAL. 31
 SQ SEQUENCE 580 AA; 63948 MW; 7CD61EAD3FDA0993 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 580;
 Best Local Similarity 28.3%; Pred. No. 1.3e-05;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSNTVYHKSEAG---TSSVFYKKTGMDLPEDTTHVFMFLINNKESYSVKDIT 74
 DB 152 TATATGRLTEGVTNETGQIENDYFFFKVGLAGE-SNOQWFMFLNVLNKSVDYEDIS 210
 OY 75 IKDQIOGGQQLDLSTLINVYTHSNYSQSAITDPEKAFPSKITYDNTKNTIDVTIP 134
 DB 211 IADROSGGQQLNKESFTFIVNDKETRY--ISLAEFEQGYGKIDFV--TDNDFMLRFY 265
 OY 135 OGGSYNSFSINKTKIT--NEQOKFEVNNNSQAWYOEHEKKEEYNGKSFNHTVHNANAG 192
 DB 266 ROKARFTSFIVRTSTITTAGOHQATFENSYDINTYOLNODATNEKNTSOV-----KNVF 320
 OY 193 IEGTVKGLKV 203
 DB 321 VEGEASGNQNV 331

RESULT 10
 OYF861 PRELIMINARY; PRT; 580 AA.
 AC OYF861:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.

OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JH2-2;
 RX MEDLINE=20407335; PubMed=10948146;
 RA Nallapareddy S.R., Singh K.V., Duh R.W., Weinstein G.M., Murray B.E.;
 RT "Diversity of ace, a gene encoding a microbial surface component
 RT recognizing adhesive matrix molecules, from different strains of
 RT enterococcus faecalis and evidence for production of ace during human
 RT infections.";
 RL Infect. Immun. 68:5210-5217(2000).
 DR EMBL: AF260879; AAC23937.1;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor.1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Signal.
 FT SIGNAL 1 31 POTENTIAL.
 SQ SEQUENCE 580 AA; 64176 MW; 060EFE50D61A5271 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 580;

Best Local Similarity 28.3%; Pred. No. 1.3e-05; Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSSTVNTVHKSEAG---TSSVFYKTKGMLPEDTTHVFMFLINNEKSYVSKDIT 74
 DB 152 TATATQRLTIEGVTNTETGQIERDYPFYKVGDLAGE-SNOVWFMFLVNLNKSQVTEEDIS 210
 OY 75 IKDQIGGQQLDLSTLINVNTGTHSNYSQSAITDEKAFPGSKITVDNTKNTIDVTIP 134
 DB 211 IADROGSGQQLNKSEFTFIVNDKETRY---ISLAEEOGGYGRKIDFV--TDNDFNLFY 265
 OY 135 OGYSYNSFESINVKTKT--NEOQKEFNNSQAMYOHGKEEVNGKSFNTHVHNANAG 192
 DB 266 RDKARFTSFIVRTSTITTEGQHQAATFENSIDINYQLNODATNEKNTSOV-----KNVF 320
 DB 193 IEGTVKGEKLV 203
 DB 321 VEGEASGNQNV 331

RESULT 11

ID 09F864 PRELIMINARY; PRT; 627 AA.
 AC 09F864;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MC02152/TX0024;
 RX MEDLINE=20407335; PubMed=10948146;
 RA Nallapareddy S.R., Singh K.V., Duh R.W., Weinstein G.M., Murray B.E.;
 RT "Diversity of ace, a gene encoding a microbial surface component
 RT recognizing adhesive matrix molecules, from different strains of
 RT enterococcus faecalis and evidence for production of ace during human
 RT infections.";
 RL Infect. Immun. 68:5210-5217(2000).
 DR EMBL: AF260876; AAC23934.1;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor.1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Signal.
 FT SIGNAL 1 31 POTENTIAL.
 SQ SEQUENCE 627 AA; 68849 MW; C9B368AE30858359 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 627;
 Best Local Similarity 28.3%; Pred. No. 1.5e-05;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;
 OY 18 TTSGNKSSTVNTVHKSEAG---TSSVFYKTKGMLPEDTTHVFMFLINNEKSYVSKDIT 74
 DB 152 TATATQRLTIEGVTNTETGQIERDYPFYKVGDLAGE-SNOVWFMFLVNLNKSQVTEEDIS 210
 OY 75 IKDQIGGQQLDLSTLINVNTGTHSNYSQSAITDEKAFPGSKITVDNTKNTIDVTIP 134
 DB 211 IADROGSGQQLNKSEFTFIVNDKETRY---ISLAEEOGGYGRKIDFV--TDNDFNLFY 265
 OY 135 OGYSYNSFESINVKTKT--NEOQKEFNNSQAMYOHGKEEVNGKSFNTHVHNANAG 192
 DB 266 RDKARFTSFIVRTSTITTEGQHQAATFENSIDINYQLNODATNEKNTSOV-----KNVF 320
 OY 193 IEGTVKGEKLV 203
 DB 321 VEGEASGNQNV 331

RESULT 12

ID 09F865 PRELIMINARY; PRT; 674 AA.
 AC 09F865;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SE33/TX1329;
 RX MEDLINE=20407335; PubMed=10948146;
 RA Nallapareddy S.R., Singh K.V., Duh R.W., Weinstein G.M., Murray B.E.;
 RT "Diversity of ace, a gene encoding a microbial surface component
 RT recognizing adhesive matrix molecules, from different strains of
 RT enterococcus faecalis and evidence for production of ace during human
 RT infections.";
 RL Infect. Immun. 68:5210-5217(2000).
 DR EMBL: AF260875; AAC23933.1;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor.1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Signal.
 FT SIGNAL 1 31 POTENTIAL.
 SQ SEQUENCE 674 AA; 73815 MW; AFD039BED47C356 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 674;
 Best Local Similarity 28.3%; Pred. No. 1.6e-05;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSSTVNTVHKSEAG---TSSVFYKTKGMLPEDTTHVFMFLINNEKSYVSKDIT 74
 DB 152 TATATQRLTIEGVTNTETGQIERDYPFYKVGDLAGE-SNOVWFMFLVNLNKSQVTEEDIS 210
 OY 75 IKDQIGGQQLDLSTLINVNTGTHSNYSQSAITDEKAFPGSKITVDNTKNTIDVTIP 134
 DB 211 IADROGSGQQLNKSEFTFIVNDKETRY---ISLAEEOGGYGRKIDFV--TDNDFNLFY 265
 OY 135 OGYSYNSFESINVKTKT--NEOQKEFNNSQAMYOHGKEEVNGKSFNTHVHNANAG 192
 DB 266 RDKARFTSFIVRTSTITTEGQHQAATFENSIDINYQLNODATNEKNTSOV-----KNVF 320
 OY 193 IEGTVKGEKLV 203
 DB 321 VEGEASGNQNV 331

RESULT 13
ID 09F858 PRELIMINARY; PRT: 319 AA.
AC 09F858:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BE83/TX0855;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapereddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260891; AAG23949.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 319 AA; 36096 MW; 7716569F98CA593C CRC64;

Query Match 15.8%; Score 177.5; DB 2; Length 319;
Best Local Similarity 25.0%; Pred. No. 7.1e-06;
Matches 53; Conservative 39; Mismatches 83; Indels 37; Gaps 7;

QY 16 VATTISGKSTNVTVHKSSEAG---TSSVFYKKTGDMLEPDTTHVRFMLINNEKSYVSKDIT 53
DB 93 IITLITNSQPNVT--ETDFGTATATQRLTIEGNTTEFGQIEODYFFFKYGDLDGE-S 149
QY 54 THVRWFLINNEKSYVSKDITTKDOIGGOQLDLSTLININVTGTHSNYSGOSAITDPER 113
DB 150 NGVRWFLINNLKNSDVTEDISADRGSGQOLNKSEFTFDIVDKETKY---ISLAEFQ 206
QY 114 APPGSKITVDNKTNTIDVTIPQGYGSYNSFSINTKKIT--NEQKEFVNNSQAWYOEHG 171
DB 207 OGYGKITYF--TDNDFNLFFYRKARFTSFIVKYSTITEAGQHATFENSYDINQNLN 264
QY 172 KEEVNGKSFNHTVHNINANAGIEGTVKGEIKY 203
DB 265 QDATNEKNTSQV-----KNVFEVGEASGNQNV 291

RESULT 14
ID 09F860 PRELIMINARY; PRT: 296 AA.
AC 09F860:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WH245/TX0635;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapereddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";

RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260887; AAG23945.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 296 AA; 33331 MW; 8D9630648275715F CRC64;

Query Match 15.6%; Score 175; DB 2; Length 296;
Best Local Similarity 27.7%; Pred. No. 1e-05;
Matches 53; Conservative 34; Mismatches 88; Indels 16; Gaps 6;

QY 18 TITSGKSTNVTVHKSSEAG---TSSVFYKKTGDMLEPDTTHVRFMLINNEKSYVSKDIT 74
DB 89 TATATQRLTIEGVTNTEGQIERDYFFFKYGDLDGE--SNQVRWFLINNLKNSDVTEDIS 147
QY 75 IKDOIGGOQLDLSTLININVTGTHSNYSGOSAITDPERKAFPGSKITVDNKTNTIDVTIP 134
DB 148 IADRGSGQOLNKSEFTFDIVDKETKY---ISLAEFQOGYKIDFV--TDNDFNLRY 202
QY 135 OGYSYNSFSINTKKIT--NEQKEFVNNSQAWYOEHGKEEVNGKSFNHTVHNINANG 192
DB 203 RDKARTSFIVKYSTITEAGQHATFENSYDINQNLNODATNEKNTSQV-----KNVF 257
QY 193 IEGTVKGEIKY 203
DB 258 VDEASGNQNV 268

RESULT 15
ID 09E080 PRELIMINARY; PRT: 319 AA.
AC 09E080:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BE88/TX0860, HG6280/TX0630, HG10528/TX0631, AND HG9829/TX0633;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapereddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260892; AAG23950.1; -.
DR EMBL; AF260884; AAG23942.1; -.
DR EMBL; AF260885; AAG23943.1; -.
DR EMBL; AF260886; AAG23944.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 319 AA; 35942 MW; 296BCB542350DA8A CRC64;

Query Match 15.6%; Score 175; DB 2; Length 319;
Best Local Similarity 27.7%; Pred. No. 1.1e-05;
Matches 53; Conservative 34; Mismatches 88; Indels 16; Gaps 6;

QY 18 TITSGKSTNVTVHKSSEAG---TSSVFYKKTGDMLEPDTTHVRFMLINNEKSYVSKDIT 74
DB 112 TATATQRLTIEGVTNTEGQIERDYFFFKYGDLDGE--SNQVRWFLINNLKNSDVTEDIS 170
QY 75 IKDOIGGOQLDLSTLININVTGTHSNYSGOSAITDPERKAFPGSKITVDNKTNTIDVTIP 134
DB 171 IADRGSGQOLNKSEFTFDIVDKETKY---ISLAEFQOGYKIDFV--TDNDFNLRY 225
QY 135 OGYSYNSFSINTKKIT--NEQKEFVNNSQAWYOEHGKEEVNGKSFNHTVHNINANG 192

Db	226	RDKARFTSFIVRTSTTTEAGQHQAPEKNSYDINTQNLNODATNEKNTSQV-----KNVF	280
Qy	193	IEGTWKGELKV	203
Db	281	VDGEASGNCNV	291

Search completed: January 30, 2002, 00:10:29
Job time: 557 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 22:10:31 ; Search time 167.16 Seconds
(without alignments)
4354.326 Million cell updates/sec

Title: US-09-813-820-3

Perfect score: 849
Sequence: 1 GACGATTAATAATGAAAAAT.....AACAGCTAATACCAAG 849

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.GeneSeq.1101.*
1: /SID88/gcgdata/geneSeq/geneSeq/NA1980.DAT.*
2: /SID88/gcgdata/geneSeq/geneSeq/NA1981.DAT.*
3: /SID88/gcgdata/geneSeq/geneSeq/NA1982.DAT.*
4: /SID88/gcgdata/geneSeq/geneSeq/NA1983.DAT.*
5: /SID88/gcgdata/geneSeq/geneSeq/NA1984.DAT.*
6: /SID88/gcgdata/geneSeq/geneSeq/NA1985.DAT.*
7: /SID88/gcgdata/geneSeq/geneSeq/NA1986.DAT.*
8: /SID88/gcgdata/geneSeq/geneSeq/NA1987.DAT.*
9: /SID88/gcgdata/geneSeq/geneSeq/NA1988.DAT.*
10: /SID88/gcgdata/geneSeq/geneSeq/NA1989.DAT.*
11: /SID88/gcgdata/geneSeq/geneSeq/NA1990.DAT.*
12: /SID88/gcgdata/geneSeq/geneSeq/NA1991.DAT.*
13: /SID88/gcgdata/geneSeq/geneSeq/NA1992.DAT.*
14: /SID88/gcgdata/geneSeq/geneSeq/NA1993.DAT.*
15: /SID88/gcgdata/geneSeq/geneSeq/NA1994.DAT.*
16: /SID88/gcgdata/geneSeq/geneSeq/NA1995.DAT.*
17: /SID88/gcgdata/geneSeq/geneSeq/NA1996.DAT.*
18: /SID88/gcgdata/geneSeq/geneSeq/NA1997.DAT.*
19: /SID88/gcgdata/geneSeq/geneSeq/NA1998.DAT.*
20: /SID88/gcgdata/geneSeq/geneSeq/NA1999.DAT.*
21: /SID88/gcgdata/geneSeq/geneSeq/NA2000.DAT.*
22: /SID88/gcgdata/geneSeq/geneSeq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	849	100.0	849	19	AAAT93437
2	849	100.0	1500	19	AAAT93438
3	790.6	93.1	4612	13	AAO24123
4	441	51.9	441	19	AAAT93436
5	170	20.0	936	22	AAAF58252
6	170	20.0	936	22	AAAF58254
7	170	20.0	936	22	AAAF58257
8	170	20.0	936	22	AAAF58259
9	170	20.0	936	22	AAAF58262
10	170	20.0	938	22	AAAF58255
11	162.4	19.1	936	22	AAAF58252

12	162.4	19.1	936	22	AAAF58254	Oligonucleotide D1
13	162.4	19.1	936	22	AAAF58257	Oligonucleotide D1
14	162.4	19.1	936	22	AAAF58259	Oligonucleotide D2
15	162.4	19.1	936	22	AAAF58262	Oligonucleotide D2
16	162.4	19.1	938	22	AAAF58255	Oligonucleotide D1
17	162.4	7.1	3869	20	AAAT13327	Enterococcus faeca
18	58.8	6.9	244	22	AAAF58238	Oligonucleotide D1
19	58.6	6.9	244	22	AAAF58238	Oligonucleotide D1
20	57	6.7	1374	20	AAAC83828	Collagen binding A
21	51	6.0	6644	20	AAAT33181	Base sequence of t
22	51	6.0	7372	20	AAAT33182	Base sequence of t
23	51	6.0	7797	20	AAAT33180	Cowpox virus bsr f
24	51	6.0	7996	20	AAAT33184	Base sequence of t
25	50.8	6.0	1524	21	AAAT70107	Plasmodium falci
26	50.6	6.0	15016	20	AAAT99560	Nucleic acid seque
27	49.4	5.8	3837	21	AAAT70211	Plasmodium falci
28	49.4	5.8	6033	21	AAAT70152	Plasmodium falci
29	49	5.8	3159	22	AAAC93000	Candida albicans C
30	48.4	5.7	1677	21	AAAT70207	Plasmodium falci
31	48.4	5.7	1865	21	AAAT70096	Plasmodium falci
32	47.6	5.6	5454	21	AAAT70236	Plasmodium falci
33	47.4	5.6	19124	18	AAAT72882	Plasmodium var-7 g
34	47.4	5.6	19124	18	AAAT72882	Plasmodium var-7 g
35	47	5.5	3567	21	AAAT70117	Plasmodium falci
36	46.8	5.5	1308	20	AAAT99503	Nucleic acid seque
37	46.8	5.5	1683	21	AAAT70112	Plasmodium falci
38	46.6	5.5	4041	21	AAAT70120	Plasmodium falci
39	46.4	5.5	1998	21	AAAT70212	Plasmodium falci
40	46.2	5.4	7458	21	AAAT70106	Plasmodium falci
41	46	5.4	3095	11	AAAT03875	Sequence encoding
42	46	5.4	5940	21	AAAT70105	Plasmodium falci
43	46	5.4	7491	17	AAAT70156	Plasmodium falci
44	46	5.4	9789	17	AAAT41852	cDNA encoding plas
45	45.8	5.4	4590	22	AAAT24065	Yeast AOD9604-asso

ALIGNMENTS

RESULT 1	AAAT93437	standard: DNA; 849 BP.
ID	AAAT93437	
AC	AAAT93437	
XX		
DT	21-MAY-1998	(first entry)
XX		
DE	Collagen binding protein M31 epitope DNA.	
XX		
KW	Collagen binding protein: cna gene; sepsis; infection;	
KW	microbial surface component regionisng adhesive matrix molecule;	
KW	MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;	
KW	epitope M31; ss.	
XX		
OS	Staphylococcus aureus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	253..849
XX		/*tag= a
XX		
PN	WO9743314-A2.	
XX		
PD	20-NOV-1997.	
XX		
PF	14-MAY-1997;	97WO-US08210.
XX		
PR	16-MAY-1996;	96US-0017678.
XX		
PA	(UABR-) UAB RES FOUND.	
PA	(TEXA) UNIV TEXAS A & M SYSTEM.	
XX		
PI	Hook M, House-Pompeo K, Patil JM, Sthanam N, Symersky J;	
XX		

DR MPI: 1998-008801/01.
 XX P-PSDB; AAM31553.
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 PS Claim 40; Page 114-115; 143pp; English.
 XX
 XX This nucleic acid comprises a portion of the Staphylococcus aureus
 CC cna gene that codes for collagen binding protein (CBP) epitope M31
 CC (see AAM31553), i.e. amino acids 61-343 of full-length CBP. Claimed
 CC 441, 849 and 1500 bp nucleic acid sequences (see AAT93436-38)
 CC respectively encode CBP epitopes M17, M31 and M55 (see AAM31552-54)
 CC that confer protection against S. aureus infection. These nucleic
 CC acid sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. These DNA
 CC segments and the peptides encoded by them are also contemplated for
 CC use in the preparation of vaccines and as carrier proteins in
 CC vaccine formulations, as well as in the formulation of compositions
 CC for the prevention of S. aureus infection.
 CC
 XX Sequence 849 BP: 348 A; 110 C; 172 G; 219 T; 0 other;
 SQ
 Query Match 100.0%; Score 849; DB 19; Length 849;
 Best Local Similarity 100.0%; Pred. No. 9.8e-177;
 Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GACGATAAATGGAATAATACAAATGTCATTTAAAGTGGCATGGCCGACAAAC 60
 Db 1 gacgataaaatggaaaaatacaaaatggtgacatgataaaatggtcgcgcaaaagc 60
 QY 61 GGTACAGTAAAGATAGAGGTTATAGTAAACAGTTCATTAAGTGAAGACG 120
 Db 61 ggtacagtaaaagatagaggttataagtaaaacagtcaccattacgttaaaagttgacag 120
 QY 121 GTGGGTCAAGCAGTATTACACAGCGGTGCACAAATTACATTCAATGATTAAGTAA 180
 Db 121 gtgggtcaagcagttattaccacagacggtgcacaaattacattcaatgataaagtagaa 180
 QY 181 AATTAAGTATGATCTTCCGGATTTCGACAATTGGAAGTCAAGGAAGAATTAAACGCCAA 240
 Db 181 aaattaagtagatcttcggatttcgacaattggaagtaagaagaatttaacgcaa 240
 QY 241 ACAATTAATCTTCAGATGACAAAGTAGCTACGATTAACATCTGGGAATTAATCAAGAAATGTT 300
 Db 241 acaataactctcagatgacaaagtagctacgataacatctcgggaataatacaacgaatglt 300
 QY 301 ACGGTTTCATTAAGTGAAGCGGGAACAGTAGTGTGTTTCTATTATTAACAGGAGATTAAG 360
 Db 301 acggttcataaaagtgaaagcgggaacagtagtcttctattataaaacggagtagatg 360
 QY 361 GTACCAAGAAAGATACGACATGTACGATGTTTAAATTAATTAACAAATGAAATTAAT 420
 Db 361 gtaccaagaagatgacgacatgtagatggttttaataataaacaagtaaaagtattat 420
 QY 421 GTATGAAAGATATTACTATTAAAGATCGATTCAAGTGGAGCAGCAGATTAAGTAAAC 480
 Db 421 gtatgaaagatattactattaaagatcagatcaagtggtgacagcagtagtaattaaagc 480
 QY 481 ACATTAACATTAAATGTGACAGGTACACATACCAATTAATTAATGTCGACAAAGTGAAT 540
 Db 481 acatttaaacatttaagtgcaggtacacatagcaattattatagtgacaaagtgaatt 540
 QY 541 ACTGATTTTGAAGAAAGCCTTCCAGGTTCTAAATTAACGTTGATTAATCGAAGAACAA 600
 Db 541 actgatattgaaagaaagccttccaggttcttaataataacgttgataatacgaagaacaa 600

Db 541 actgatattgaaagaaagccttccaggttcttaataataacgttgataatacgaagaacaa 600
 QY 601 ATTGATGTAAACATTTCCACAGGCTATGGGTCAATATATATGTTTTCAATTAACATCAAA 660
 Db 601 attgatgttaaacatttcacaaagcgtatgggtcatatataagttttcaatttaactacaa 660
 QY 661 ACCAAATTTACGAATGACAGCAAAAAAGGTTTGTTAAATTAATTCACAACTTGATCA 720
 Db 661 accaaattacgaatgaaagcaaaaaaggtttgtaataatcacaaagcttgatcaaa 720
 QY 721 GAGCATGTGAAGAAAGATGGAACGGGAATCATTTAATCATCTGTGCACATATTAA 780
 Db 721 gagcatgtgaagaagaagtgaaacgggaatcaattacatcacgtgcacatatattat 780
 QY 781 GCTAATGCCGATTATGTAAGTACTGTAAAGGTGAATTAAGTTTAAACAGCATAAA 840
 Db 781 gctaatgccggtatgtaaggtactgttaaaagtgtaattaaagtttaaacagcataaa 840
 QY 841 GATACCAAG 849
 Db 841 gataccaag 849
 RESULT 2
 AAT93438
 ID AAT93438 standard; DNA; 1500 BP.
 XX
 AC AAT93438;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE Collagen binding protein M55 epitope DNA.
 XX
 DE Collagen binding protein; cna gene; sepsis; infection;
 KW microbial surface component recognising adhesive matrix molecule;
 KW MSCRAMM; adhesion; vaccine; immunisation; diagnosis; therapy;
 KW epitope M55; ss.
 XX
 OS Staphylococcus aureus.
 XX
 PN MO9743314-42.
 XX
 PD 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US08210.
 XX
 PR 16-MAY-1996; 96US-0017678.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (TEXA) UNIV TEXAS A & M SYSTEM.
 XX
 PI Hook M, House-Pompeo K, Patil JM, Sthanam N, Symersky J;
 XX
 DR MPI: 1998-008801/01.
 XX
 DR P-PSDB; AAM31554.
 XX
 XX Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 PS Claim 40; Page 116-117; 143pp; English.
 XX
 XX This nucleic acid comprises a portion of the Staphylococcus aureus
 CC cna gene that codes for collagen binding protein (CBP) epitope M55
 CC (see AAM31554), i.e. amino acids 30-531 of full-length CBP. Claimed
 CC 441, 849 and 1500 bp nucleic acid sequences (see AAT93436-38)
 CC respectively encode CBP epitopes M17, M31 and M55 (see AAM31552-54)
 CC that confer protection against S. aureus infection. These nucleic
 CC acid sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in

CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. These DNA
 CC segments and the peptides encoded by them are also contemplated for
 CC use in the preparation of vaccines and as carrier proteins in
 CC vaccine formulations, as well as in the formulation of compositions
 CC for the prevention of *S. aureus* infection.

XX
 Sequence 1500 BP; 628 A; 198 C; 296 G; 378 T; 0 other;

Query Match 100.0%; Score 849; DB 19; Length 1500;
 Best Local Similarity 100.0%; Pred. No. 1.1e-176;

Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACGATAAATAATGAAAAATACAAAATGGTACATGATTTAAAGTCGACGCCGACACAGC 60
 DB 94 gacgataaataatgaaaaatacacaatggtgacatgattaaagtgccatgagccgacaagc 153
 QY 61 GGTACGATAAAGATAGAGGTTATGATTAACATCCATTACGTGTTAAAGGTGAACAG 120
 DB 154 ggtacagtaagatagaggttataagtaaacagtaaccatlaacgtgtaaggtgacag 213
 QY 121 GTGGGTCAAGCAGTTATACACGACAGCGTGCAACAATTACATTCATGATAAGTAGAA 180
 DB 214 gtgggtcaagcagttatacaccagacggtgacacaattacatcaatgataagtagaa 273
 QY 181 AATTAAGTGATGTTTGGGATTTCGAGAAATTTGAATGACAGGAAATTTTAACGCAA 240
 DB 274 aatlaagtgatgtttccgagatttcgagaatttgaagtaacaagaaatttaacgcaa 333
 QY 241 ACAAAATCTTCAGATGACAAAGTACGATACATCTGGGAATTAATCAACGAATGTT 300
 DB 334 acaaaatcttcagatgacaaagtagtaacgatacaatcgtggaataaatacaacgaatgtt 393
 QY 301 ACGGTTCAATAAAGTGAAGACGCGAACAAGTAGAGTTTTCATATTAATGAACGAGATATG 360
 DB 394 acggttcataaagtgaaagcggaacaagtagtgcattcattataaaccggaagatag 453
 QY 361 CTACCGAAGATACGACACATGTACAGTGGTTTAAATTTAACAATGAAAAAGTTAT 420
 DB 454 ctaccgaagatatacgacacatgacatgtgttttaataataatacgaatgaaaaagttat 513
 QY 421 GATTCGAAGATTTACTTAAGATGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 480
 DB 514 gatacgaagataatactataaagatcagatcagatcagatcagatcagatcagatcagat 573
 QY 481 ACATTAACATTAATGATGACAGTACATACGATTAATTAATTAATTAATTAATTAATTA 540
 DB 574 acatataacatataatgacaggtacacatagcaatataatgacagaaagtgcaatt 633
 QY 541 ACTGATTTTGAATAAGCCCTTTCAGGTTCTAAATAACTGTTGATTAATGAAGAACA 600
 DB 634 actgattttgaaaaagcccttcaggttctaataactgttgataatacgaagaacaca 693
 QY 601 ATTGATGATTAACAATTCCACAAGGCTATGGTCTATTAATTAATTAATTAATTAATTA 660
 DB 694 attgatgataacaattccacaaggtctatgggtcattataatagtttccaactaactaaca 753
 QY 661 ACCAAATTTACGAATGACAGCAAAAAGTTGTTAAATTAATTAATTAATTAATTAATTA 720
 DB 754 accaaataatgaatgaacagcaaaaaagtttctgtaataatacacaagcttctgtatcaa 813
 QY 721 GAGCATGATTAAGGAGAAGTGAACGGAATCAATTAATCAATCAATCAATTAATTAAT 780
 DB 814 gagcatgataaggaagaagtgaaacgggaataatcaatcaatcaatctgtcaacaatlaa 873
 QY 781 GCTAATGCCGGTATTGAAGTACTGTAAGGATTAAGGATTAAGGATTTTAAACGAGTAAA 840
 DB 874 gctaatagccggtattgaagtaactgttaaaagtgaaatlaaaagtttaaaacaggtataa 933
 QY 841 GATACCAAG 849
 |||||||

DB 934 gataccaag 942
 RESULT 3
 ID AAQ24123
 XX AAQ24123 standard; DNA; 4612 BP.
 AC AAQ24123;
 XX
 XX 04-DEC-1992 (first entry)
 DE p16 and cCOLR6a assembled cbp sequence.
 XX CBP; collagen binding protein; ss.
 KW
 XX Staphylococcus aureus.
 OS
 XX
 XX Key
 FH misc_RNA
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "Claim 6; page 26-28"
 FT 1..8
 FT /tag= b
 FT /note= "part of the vector MCS"
 FT 9..3930
 FT /tag= c
 FT /note= "Insert in p16"
 FT 2385..4612
 FT /tag= d
 FT /label=
 FT /note= "5' end of Insert in cCOLR6a"
 FT 917..924
 FT /tag= e
 FT 931..4488
 FT /tag= f
 FT /label= CBP
 FT 931..1017
 FT /tag= g
 FT /label= S
 FT /note= "signal sequence"
 FT 1018..2532
 FT /tag= h
 FT /label= A
 FT /note= "see CC"
 FT 2533..3093
 FT /tag= i
 FT /label= B1
 FT /note= "see CC"
 FT 3094..3654
 FT /tag= j
 FT /label= B2
 FT /note= "see CC"
 FT 3655..4215
 FT /tag= k
 FT /label= B3
 FT /note= "see CC"
 FT 4216..4407
 FT /tag= l
 FT /label= W
 FT /note= "see CC"
 FT 4408..4467
 FT /tag= m
 FT /label= M
 FT /note= "see CC"
 FT 4468..4485
 FT /tag= n
 FT /note= "charged C-terminal"
 PF 22-OCT-1991; 91WO-SR00207.

XX 22-OCT-1990: 90SE-0003374.
 PR (ALFA) ALFA IAVAI AGRI INT AB.
 PA
 XX
 PI Guss BM, Hook M, Jonsson H, Lindberg KM, Patti J, Signaes LC;
 PI Switalski LM;
 DR WPI; 1992-16709/20.
 DR P-PSDB; AAR2675.
 XX
 PT Hybrid DNA molecule encoding S. aureus collagen binding protein -
 PT protein is expressed in E.coli and used for diagnosis e.g. of
 PT septic arthritis
 XX
 PS Disclosure: Fig 2; 40pp; English.
 CC The amino acid encoded by this sequence (assembled from partially
 CC homologous p16 and cCOL6A inserts) comprises a structure resembling
 CC a signal sequence. Following this region, a region called A (tag- h)
 CC is found followed by a repetitive stretch of 187 amino acids called B1
 CC (tag- i), B2 (tag- j) and B3 (tag- k). Directly following these
 CC regions there is a region called W (tag- l) which consists of a
 CC repetitive, hydrophilic structure contg. several proline residues.
 CC This region is thought to mediate the binding of the protein to the
 CC cell wall. The amino acid sequence nearest to the C-terminal end
 CC consists of a long stretch of hydrophobic residues followed by some
 CC charged amino acids (tag- n). This region is called M (tag- m).
 XX
 SQ Sequence 4612 BP; 1946 A; 675 C; 869 G; 1122 T; 0 other;

Query Match 93.1%; Score 790.6; DB 13; Length 4612;
 Best Local Similarity 97.7%; Pred. No. 7.3e-164;

Matches 835; Conservative 0; Mismatches 14; Indels 6; Gaps 3;

QY 1 GACGATAAATGGAATAATACAAATGTCATGATTAAGTCGACGCCGACACAGC 60
 DB 1111 gacgataaataatggaataatgacaaatgacatgataaagtcgacacagc 1170
 QY - 61 GGTACAGTAAAGTATAGAGGTTTACTTAAACAGTACCACTTAAAGTGAACAG 120
 DB 1171 ggtacagtaagataagaggtatagtaataaagcattacatcagcttaagtgacag 1230
 QY 121 GTGGGTCAAGCAGTATATACACGAGCGTGCACAAATTACATTCATGATTAAGAA 180
 DB 1231 gtgggtcaagcagttatcaccagcagtgacaaattacatcattgataagtagaa 1290
 QY 181 AATTATAGTATGTTTGGGATTTGCAGAAATTTGAATGACAGGAAATTTAACGCA 240
 DB 1291 aattatagatgtttcgggatttgcagaatttgaagtaagaagaatttaacgcaa 1350
 QY 241 ACAAAATCTTCGATGACAAAGTAGTACGATCATCTGGGAAATTAACAAGATGTT 300
 DB 1351 acaaaatcttcgatgacaaagtagtacctgacgataacatcgggaaataacagaaatgtt 1410
 QY 301 A--CGGTTCTAAAGTGAAGCGGGAACAAGTAGTCTTTTCTATTATAA---CGGAG 355
 DB 1411 atcggttgataaagaatgagcggaacagtagtcttctaataataaagcgggaag 1470
 QY 356 AATATGCTACC-AGACATACGACACATGATCGATGTTTAAATATTAACAATGAAAA 414
 DB 1471 aatgctaccagaagaatgacacatgacatgtcttctaataataacagtaaaaa 1530
 QY 415 AGTTATGTTTCGAAAGATTTTACTATAAGATCAGATTCAGAGTGACAGCATAGAT 474
 DB 1531 agttaatgctcgaagaatcattactataaagatgagattcaagtgagacgcttagat 1590
 QY 475 TTAAGCATTAAACATTTATGACAGGTACACATAGCAATATATATAGTGACAAAGT 534
 DB 1591 ttaagcattaaacattatgtagacagtgacacatagcaattatattatgtgacaaagt 1650
 QY 535 GCAATTACTGATTTGAAAAAGCCTTCCAGGTTCTTAAATAAAGTGTATATACGAAG 594

DB 1651 gcaattactgatttgaaaaaagccttcacaggttcctaaataactgtgataacgaag 1710
 QY 595 AACCAATGATGATTAACATTCACAAAGGOTATGGTCATATATATGTTTTCATTAAC 654
 DB 1711 aaccaatgatagttaacaaatccacaaggtatggtcatataatagtttccaattac 1770
 QY 655 TACMAAACCAAAATTAAGATGAATGACGACAAAAGAGTTTGTATTAATTCACAAAGCTTG 714
 DB 1771 tacaacaacaaattacgatagtgacagcaaaaagagttgttaataatccaagcttg 1830
 QY 715 TATCAAGAGCATGCTAAGAGAGAGTGAACGGGAATTCATTTATATACTGTGCAAT 774
 DB 1831 tatcaagagcatgttaaggaagaagtgaaaggaatcattatcatctgtgcacaa 1890
 QY 775 ATTAATGCTTAATGCGGTATTTGAAGTACGTATTAAGGTAATTAAGTTTAAACAG 834
 DB 1891 attaatgcttaatgcggtatltgaaggtactgttaaaagtgaaatlaaaagtltaaaacag 1950
 QY 835 GATTAAGATATACCAAG 849
 DB 1951 gataaagataccaa 1965

RESULT 4
 ID AAT93436 standard; DNA; 441 BP.
 AC AAT93436;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE Collagen binding protein M17 epitope DNA.
 KW Collagen binding protein; cna gene; sepsis; infection;
 KW microbial surface component recognising adhesive matrix molecule;
 KW MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;
 KW epitope M17; ss.
 OS Staphylococcus aureus.
 XX
 PN WO9743314-A2.
 XX
 PD 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US08210.
 XX
 PR 16-MAY-1996; 96US-0017678.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (TEXA) UNIV TEXAS A & M SYSTEM.
 XX
 PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 XX
 DR WPI; 1998-008801/01.
 XX
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 40; Page 113; 143pp; English.
 CC This nucleic acid comprises a portion of the Staphylococcus aureus
 CC cna gene that codes for collagen binding protein (CBP) epitope M17
 CC (see AAW31552). i.e. amino acids 151-297 of full-length CBP. Claimed
 CC 441, 849 and 1500 bp nucleic acid sequences (see AAT93436-38)
 CC that confer protection against S. aureus infection. These nucleic
 CC acid sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in

CC screening, diagnostic and therapeutic applications including active
CC and passive immunisation and methods for the prevention of
CC bacterial colonisation in an animal such as a human. These DNA
CC segments and the peptides encoded by them are also contemplated for
CC use in the preparation of vaccines and as carrier proteins in
CC vaccine formulations, as well as in the formulation of compositions
CC for the prevention of *S. aureus* infection.

SQ Sequence 441 BP; 180 A; 60 C; 78 G; 123 T; 0 other;

Query Match	51.98; Score 441; DB 19; Length 441;
-------------	--------------------------------------

```

Best Local Similarity 100.0%; Pred. No. 1.2e-8;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	271	ATACATCTGGGAATTAATCAAGCAATGTTACGGTTCATATAAGTGAAGCGGAAACAAT	330
Db	1	atacacctcgggaaataacacgaacglttaaggltcataaagtgaaagcggaacaagt	60
QY	331	AGTGTGTTCTATTATTAACGCGGAGATATGCTACACAGAAATACGACATGTACGATGG	390
Db	61	agtgcttccatctatctaaacgggagatgctaccagaagaatacgacacatgtacggtg	120
QY	391	TTTTTAATATTAACATGAAAAAGCTTAGTATCGAAATATTTACTATTAAGGATCAG	450
Db	121	ttttaataatlaacaatgaaaaaglttaagltcgaagaatatatactataaaggatcag	180
QY	451	ATTCAAGGTGACGACGAGTTAGATTTAGACATTAACATTAATATGACAGGTACAT	510
Db	181	attcaagtgagcaagcgatagatttaagaacacatlaaaccttaatgtgaaggtacacat	240
QY	511	AGCAATTATTATAGTGGACAAAGTGCATTTACTGATTTTGAAAAAGCCTTCCAGTTCT	570
Db	241	agcaattattatagtggaacaagtgccattactgattttbaaaaagccttccaagttct	300
QY	571	AAATTACTGTGATTATPAGAAACACACATTGATGATTAACAATTCACAAAGGTATGG	630
Db	301	aaaaaacctgttgataataacgaagaacacacatgtatgataacattccacaaggtatg	360
QY	631	TCATATATATAGTTTTCATTAATTAACACAAAACCAAAATTAAGATGAACGACAAAAGAG	690
Db	361	tcatataatagtttccaatlaacatacaaaaacaaatctcgaatgacagcaaaaagag	420
QY	691	TTTGTTAATTAATTCACAAGCT 711	
Db	421	tttgttaataatcttcaagct 441	

RESULT	5
AAF58252/c	
ID	AAF58252 standard; DNA; 936 BP.
XX	
AC	AAF58252;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1835.
XX	
KM	Electron-transfer group; ETM; mismatch; genotyping.
KM	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	W0200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;
PI

WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electroc
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match	20.08;	Score 170;	DB 22;	Length 936;
-------------	--------	------------	--------	-------------

Best Local Similarity 1.0%; Pred No. 2 le-28;
Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

[illegible]

```

QY 779 ATGCTATGCCGTATGTGAAGTACTGTAAGGCGAATTTAAAGTTTAAACAGGATA 838
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 12
QY 839 AAGATA 844
   : : : : :
Db 11 WWWWWW 6

RESULT 6
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KM Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
PI
PI -Umek RM;
XX
WPI: 2001-159728/16.
XX
DR
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
CC
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 2, 1e-28;
Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAGATAGAGGGTTATAGTAAACACTACATTAACTGTTAAAGTGAAC 118
   | | : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 GGGCGMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 732
QY 119 AGGTGGGTCAAGCAGTTATTACACGACGAGTGAACAATTACATTCATGATAAGTAG 178
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 672
QY 179 AAAAATTAAGTATGTTGGGATTTGACGAATTTGAAGTACAAGAGAAATTTAAACG 238
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 612
QY 239 AAACAATATCTCAGATGCAAGATGCTACGATACTCGGAATTAATCACAAG 298
   : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 611 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 552
QY 299 TTACGGTTCATAAAGTGAAGCGGAACAAGTAGTGTCTTCTATTATAAAGGAGATA 358
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 492
QY 359 TCGTACCAGAGATACGACACATGACGATGTTTAAATATTAACAATAAAGTT 418
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 432
QY 419 ATGTATCGAAAGATATTACTATTAAGATCAGATCAAGTGGACGACGACTGATTTA 478
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 372
QY 479 GCACATTAAACATTAAATGTGACAGTACACATAGCAATTATATAGTGAACAAGTCAA 538
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 312
QY 539 TTACTGATTTGAAAAAGCCTTCCAGGTTCTAAATTAAGTGTGATATACGACAGACA 598
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 252
QY 599 CAATGTATGTAACAATTCACAAAGCGATGCGTATATAGTTTTCATTAAGTACA 658
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 192
QY 659 AAACAAAATTTACGAATGAACAGCAAAAGAGTTGTTAATAATGACAGCTGTATC 718
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 132
QY 719 AAGACATGTTAAGGAAGAGTGAACGGCAATCTTAATCACTGTCACAAATTA 778
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 72
QY 779 ATGCTATGCCGTATGTGAAGTACTGTAAGGCGAATTTAAAGTTTAAACAGGATA 838
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 12
QY 839 AAGATA 844
   : : : : :
Db 11 WWWWWW 6

RESULT 7
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KM Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
PI
PI -Umek RM;
XX
WPI: 2001-159728/16.
XX
DR
XX

```

PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.

SO Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;
 Best Local Similarity 1.0%; Pred. No. 2.1e-28;
 Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 118
 DB 791 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 732
 QY 119 AGGTGGTCAAGAGCTTATTACACAGCGTGCAACAATTACATCAATGATTAAGTAG 178
 DB 731 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 672
 QY * 179 AAAAATTAAGTGTGTTGGGATTTGCAAGATTTGAAGTACAAGAAATTTAAACGC 238
 DB 671 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 612
 QY 239 AAACAATCTCTCAGATGACAAAGTACGATGAATCTGGGAATTAATCAAGCATG 298
 DB 611 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 552
 QY 299 TTACGGTTCATAAAGTGAAGCGGAAAGTGTGTTCTATTATTAACGGGAGATA 358
 DB 551 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 492
 QY 359 TGCTACAGAGATGACAGACATGATGTTTAAATATTAACAATGAAGAAAGTT 418
 DB 491 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 432
 QY 419 AGTATCGAAGATGATTAATAAGATGATGATCAAGTGAAGTGAAGTGAAGTGA 478
 DB 431 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 372
 QY 479 GCACATTAACATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 538
 DB 371 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 312
 QY 539 TTACTGATTTTGAAGAAAGCTTTCAGGTTCTAAATTAATGATTAATCAAGACACA 598
 DB 311 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 252
 QY 599 CAATTATGTAACATTCACAGAGCTATGGGATATTAATGATTTTCAATTAAGTACA 658
 DB 251 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 192
 QY 659 AAACCAAAATTTACGAATGAAGACAAAGAGTTGTTAATTAATTAACAAGCTTGATC 718
 DB 191 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 132
 QY 719 AAGAGCATGTAAGAGAGAGTGAAGCGAAATCATTAATCATCTGTCACAAATATTA 778
 DB 131 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 72
 QY 779 ATGCTAATGCCGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 838
 DB 71 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 12
 QY 839 AAGATA 844

DB 11 WWWWWWW 6

RESULT 8

AA58259/C

AA58259 standard; DNA; 936 BP.

AA58259;

24-APR-2001 (first entry)

Oligonucleotide D2004.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

MO200107665-A2.

26-JUL-2000; 2000MO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in

hybridization assays, e.g. for genotyping, allowing repeat analyses on

a single surface

Example 6; Page 128; 159pp; English.

The present invention relates to a composition comprising two nucleic

acids each containing an electron-transfer group (ETM) having

different redox potentials. The invention is used for electronic

detection of nucleic acids, especially of substitutions (mismatches)

and single-nucleotide polymorphisms, e.g. for genotyping,

monitoring gene expression.

Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;

Best Local Similarity 1.0%; Pred. No. 2.1e-28;

Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 118
 DB 791 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 732
 QY 119 AGGTGGTCAAGAGCTTATTACACAGCGTGCAACAATTACATCAATGATTAAGTAG 178
 DB 731 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 672
 QY 179 AAAAATTAAGTGTGTTGGGATTTGCAAGATTTGAAGTGAAGTGAAGTGAAGTGA 238
 DB 671 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 612
 QY 239 AAACAATCTCTCAGATGACAAAGTACGATGAATCTGGGAATTAATCAAGCATG 298
 DB 611 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 552
 QY 299 TTACGGTTCATAAAGTGAAGCGGAAAGTGTGTTCTATTATTAACGGGAGATA 358
 DB 551 GGGCGTAAAGATGAGGTTATAGTAAACAGTACCATTAACTGTTAAAGTGAGAC 492


```
OY 359 TGCCTACGAAGATACGACATGTACATGCTTTTAAATTTAAACAATGAAAAAGTT 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 432
OY 419 ATGTATCGAAGATATTACTATAAAGATTCAGATTCAGAGTGACAGCGTTAGATTAA 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 372
OY 479 GCACATTAACATTAATGTGACAGGTACACATAGCAATTAATATAGTGACAAAGTGCA 538
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 312
OY 539 TTACTGATTTTGAAGAACGCTTTCCAGGTTCTAAATAACTGTGATTAATACGAAGACA 598
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 252
OY 599 CAATGATGTACAAATTCACAAAGGCTATGGCTCATATAATAGTTTCAATTACTACA 658
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 192
OY 659 AAACCAAAATTAAGATGACAGCAAGCAAAAGAGTTGTTAATTAATCACAAGCTGTGATC 718
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 132
OY 719 AAGACATGTTAAGAGAAGTGAACGGGAATCATTTATCATCTGTGCACATATTA 778
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 72
OY 779 ATGCTAATGCCGTATGTGAAGTACTGTAAAGTGAATTAAGTTTAAACAGGATA 838
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 12
OY 839 AAGATA 844
      : : : :
Db - 11 WWWWXXX 6
```

RESULT 9
AAFS8262/c
ID AAF58262 standard; DNA; 936 bp.
XX
AC AAF58262;
XX
DT 24-Apr-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KM Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
XX WPI: 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
PS Example 6; Page 128; 159pp; English.

```
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 bp; 5 A; 139 C; 10 G; 6 T; 776 other;
```

Query Match 20.0%; Score 170; DB 22; Length 936;
Best Local Similarity 1.08; Pred. No. 2,1e-28;
Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

```
OY 59 GCGGTACAGTAAGATAGAGGGTTATAGTAAACAGTACCACTTAACGTAAAGTGAAC 118
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 GGGCXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 732
OY 119 AGGTGGTCAAGCAGTTATTAACACAGCGTGCAACAATTAATCAATGATTAAGTAG 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 672
OY 179 AAAAAATTAAGTATGTTTGGGATTTGCAGAATTGAATTAAGAAAGAAATTTAACCC 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 612
OY 239 AAACAAATTAATCACTAGTCAAGTACGATAAATCTGGATAAATCAACGATG 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 552
OY 299 TTACGGTTCAATAAAGTGAAGCGGAACAAGTAGTGTGTTTCTATTATAAAGCGAGATA 358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 492
OY 359 TGCCTACGAAGATACGACATGTACGATGGTTTAAATTAATCAATGAAAAAGTT 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 432
OY 419 ATGTATCGAAGATATTACTATAAAGATCAGATTCAGAGTGACAGCGATTAGATTAA 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 372
OY 479 GCACATTAACATTAATGTGACAGGTACACATAGCAATTAATTAATGACAAAGTGCA 538
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 312
OY 539 TTACTGATTTTGAAGAACGCTTTCCAGGTTCTAAATAACTGTGATTAATACGAAGACA 598
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 252
OY 599 CAATGATGTACAAATTCACAAAGGCTATGGCTCATATAATAGTTTCAATTACTACA 658
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 192
OY 659 AAACCAAAATTAAGATGACAGCAAGCAAAAGAGTTGTTAATTAATCACAAGCTGTGATC 718
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 132
OY 719 AAGACATGTTAAGAGAAGTGAACGGGAATCATTTATCATCTGTGCACATATTA 778
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 72
OY 779 ATGCTAATGCCGTATGTGAAGTACTGTAAAGTGAATTAAGTTTAAACAGGATA 838
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 12
OY 839 AAGATA 844
      : : : :
Db 11 WWWWXXX 6
```

RESULT 10

AAFS8252/c
 ID AAF58255 standard; DNA: 938 BP.
 AC AAF58255;
 DT 24-APR-2001 (first entry)
 DE Oligonucleotide D1876.
 XX
 XX Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 PN MO200107665-A2.
 XX
 XX 01-FEB-2001.
 PD
 XX 26-JUL-2000; 2000MO-US20476.
 PE
 XX 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 PA
 XX timek RM;
 PI
 DR WPI; 2001-159728/16.
 XX
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS - Example 6; Page 127; 159pp; English.
 XX
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 CC
 CC Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
 SQ

Query Match 20.0%; Score 170; DB 22; Length 938;
 Best Local Similarity 1.0%; Pred. No. 2.1e-28;
 Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAGATAGAGGTTATAGTAAACAGTACATTAAGTGAAC 118
 DB 791 GGGCGMMWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 732
 QY 119 AGGTGGTCAAGCATTTATACACAGACGCGTCAACATTAATGATGAAGTAG 178
 DB 731 WW 672
 QY 179 AAAAATTAGTATGTTTCGGGATTTGCAAGATTGAAGTACAGGAAGAAATTAAAGC 238
 DB 671 WW 612
 QY 239 AAACAATTAATTCAGATGACAAAGTACGATTAACATCGGGAATTAATCAAGCATG 298
 DB 611 WW 552
 QY 299 TTACGGTTTCAATAAGTGAAGCGGAGCAAGTATGTTTCTATTATAAACGGAGATA 358
 DB 551 WW 492
 QY 359 TGTACCAAGATAGCAGCATGTGATGTTTAAATATTAAACATGAAAAAGTT 418
 DB 491 WW 432

QY 419 ATGTATCGAAAGATATTACTATTAAGATGATTCAGGTGACAGCAGTATGATTTAA 478
 DB 431 WW 372
 QY 479 GCACATTAACATTAATGTGACAGGTACACATTAATTAATAGTGACAAAGTGCA 538
 DB 371 WW 312
 QY 539 TTACTGATTTGAAAAAGCCCTTCAGGTTCTAAATTAAGTGTGATTAATGCAAGACA 598
 DB 311 WW 252
 QY 599 CAATTGATGTACATTCACAAAGCGTATGCGTATATTAATGTTTCAATTAAGTACA 658
 DB 251 WW 192
 QY 659 AAACCAAAATTACGAATGACAGCAAGAAAGATTGTTAATAATTCACAAAGCTGTATC 718
 DB 191 WW 132
 QY 719 AAGACATGTGAAGAAAGTGAACGGAAATCATTTATCATCTGTGACATATTA 778
 DB 131 WW 72
 QY 779 ATGCTAATGCCGTATGTAAGTACTGTAAGAAGTGAATTAAGTTTAAACAGATA 838
 DB 71 WW 12
 QY 839 AAGATA 844
 DB 11 WWWWWW 6

RESULT 11
 AAF58252
 ID AAF58252 standard; DNA: 936 BP.
 AC AAF58252;
 DT 24-APR-2001 (first entry)
 DE Oligonucleotide D1835.
 XX
 XX Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 PN MO200107665-A2.
 XX
 XX 01-FEB-2001.
 PD
 XX 26-JUL-2000; 2000MO-US20476.
 PE
 XX 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 PA
 XX timek RM;
 PI
 DR WPI; 2001-159728/16.
 XX
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 19.1%; Score 162.4; DB 22; Length 936;
Best Local Similarity 0.5%; Pred. No. 9.7e-27;
Matches 4; Conservative 522; Mismatches 258; Indels 0; Gaps 0;

```
OY 63 TACAGTAAGATGAGGGTTATAGTAAACAGTACCATTAAGTGAAGTGAACAGGT 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6  www. .... 65
OY 123 GGGTCAGAGATTATTAACAGAGCGGTGCACAAATTACATTAAGTAAGTAAAGTAA 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66  www. .... 125
OY 183 ATTAAGTGTGTTGGGATTTCCAGATTGGAAGTACAGGAAGAAATTTAACGCAAC 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 www. .... 185
OY 243 AATATCTCAGATGACAAAGTACGATACATCGGGAATAATCAACGAATGTAC 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 www. .... 245
OY 303 GGTTCATAAAGTGAAGCGGGAACAAGTGTCTTCTATTATAAACGGAGATATGCT 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 www. .... 305
OY 363 ACCAGAGATACGACATGTACGATGTTTAAATATTACAAATGAAAAAGTATGT 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 www. .... 365
OY 423 ATCGAAGATATTATTAAGGATCAGATTCAAGGTGACAGACAGATTAGATTAAAC 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 www. .... 425
OY 483 ATTAACATTATGTGACAGGTACATAGCAATTTATTAGTGACAAAGTCAATTAC 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 www. .... 485
OY 543 TGTATTTGAAAAAGCCTTCCAGGTTCTAAATAACTGTTGATTAATCGAAGACACAT 602
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 www. .... 545
OY 603 TGATGTAAACATTCACAAGGCTATGTCATATAATAGTTTCAATTAACTAACAAC 662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 546 www. .... 605
OY 663 CAAATATTAGCAATGAACAGCAAAAGAGTTTGTAAATTACAAAGCTTGATACAGA 722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 www. .... 665
OY 723 GCATGTGAAGGAAGTGAACGGGAATCATTTAATCACTGTGACATTTTAATGC 782
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 666 www. .... 725
OY 783 TAATGCCGATTAATGAAGTCTGTAAGGTTGAATTAAGTTTAAACAGGATAAGA 842
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 726 www. .... 785
OY 843 TACC 846
   : : |
Db 786 wwg 789
```

RESULT 12

AAFS8254

ID AAF58254 standard; DNA; 936 BP.

XX AAF58254;

XX

DT 24-APR-2001 (first entry)

XX Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;

KM gene expression; ss.

XX Synthetic.

PN WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

DR UmeK RM;

PT Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PS a single surface - Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

XX

Query Match 19.1%; Score 162.4; DB 22; Length 936;

Best Local Similarity 0.5%; Pred. No. 9.7e-27;

Matches 4; Conservative 522; Mismatches 258; Indels 0; Gaps 0;

```
OY 63 TACAGTAAGATGAGGGTTATAGTAAACAGTACCATTAAGTGAAGTGAACAGGT 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6  www. .... 65
OY 123 GGGTCAGAGATTATTAACAGAGCGGTGCACAAATTACATTAAGTAAAGTAAAGTAA 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66  www. .... 125
OY 183 ATTAAGTGTGTTGGGATTTCCAGATTGGAAGTACAGGAAGAAATTTAAACGCAAC 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 www. .... 185
OY 243 AATATCTCAGATGACAAAGTACGATACATCGGGAATAATCAACGAATGTAC 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 www. .... 245
OY 303 GGTTCATAAAGTGAAGCGGGAACAAGTGTCTTCTATTATAAAGCGGAGATATGCT 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 www. .... 305
OY 363 ACCAGAGATACGACATGTACGATGTTTAAATATTAAACGAAGAAAGTATGT 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 www. .... 365
OY 423 ATCGAAGATATTATTAAGGATCAGATTCAAGGTGACAGACAGATTAGATTAAAC 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 www. .... 425
OY 483 ATTAACATTATGTGACAGGTACATAGCAATTTATTAGTGACAAAGTCAATTAC 542
```

Dd	426	www.....
Oy	543	TGATTTGGAAAAACCTTCCAGGTTCTAAATAACTGTGTATATACGAAGAACAAT 602
Dd	486	www.....
Oy	603	TGATGTAAACATTCCCAAGAGCATGGTCATATAATAGTTTTTCATTTAACTAACA AAC 662
Dd	546	www.....
Oy	663	CAAATTTACGAATGACAGCAGAAAAGATTGTATAATATTCACAGCGTTGGATACAGA 722
Dd	606	www.....
Oy	723	GCATGTAAAGGAAGAAGTAGAAGGAAATCATTTATCATCTGTGCACAAATTTAATGC 782
Dd	666	www.....
Oy	783	TAAATGCCGATATTGAAGTGACTGTAAAGTGATTAAGTTAAAGTTAAAAACAGATTAAGA 842
Dd	726	www.....
Oy	843	TACC 846
Dd	786	wmgc 789

RESULT_13

AAF58257 standard; DNA; 936 BP.

AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

KW electron-transfer group; ETW; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
CS
OS
PN W0200107665-A2.
PD 01-FEB-2001.
PF 26-JUL-2000; 2000MO-US20476.
PR 26-JUL-1999; 990S-0145695.
PR 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
P1 umek RM:
XX WP1: 2001-159728/16.
PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -
XX Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETW) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match	19.1%	Score 162.4	DB 22	Length 936
Best Local Similarity	0.5%	Pred. No. 9,7e-27		
Matches 4	Conservative 522	Mismatches 258	Indels 0	Gaps 0
0Y	63	TACACGTAAAGATAGAGGGTTATAGTAAACATACATTAATCTGTTAAAGTGACACAGCT	122	
Db	6	65	
0Y	123	GGGTACAGCAGTTATTACACGACGCGTCAACATTACATTCAATGATTAAGTGAATA	182	
Db	66	125	
0Y	183	ATTAAGTGATGTTGGGATTTGCAGAAATTGACAGACAGGAAGAAATTTAACGCAAC	242	
Db	126	185	
0Y	243	AAATACCTCAGATGACAAAGTAGTACGATTAACATCTGGAAATTAATCAAGAAATGTTAC	302	
Db	186	245	
0Y	303	GGTTATATAAATGGAAGCGGGAACAAGTGTGTTTCTTTTAAACGGGAGATATGCT	362	
Db	246	305	
0Y	363	ACCAGAGATACGACACATGTACGATGTTTAAATTAATTAACAATGAAGAAATGTGT	422	
Db	306	365	
0Y	423	ATCGAAGATATTACTATAAAGATCAGATTCAGAGTGACAGCAAGTATTAAGCAC	482	
Db	366	425	
0Y	483	ATTAAACATTATGTGACAGGTACACATATTTATCTGGACAAAGTGCATTAC	542	
Db	426	485	
0Y	543	TGATTTTGAAGAAAGCCTTCCAGTTCCTAAATTAATCTGTTGATATACGAAGACAAAT	602	
Db	486	545	
0Y	603	TGATGTAAACATTCCACAAAGGCTATGGTCAATATTAAGTTTTCATTAACTACAAAAC	662	
Db	546	605	
0Y	663	CAAAATTACGATGTAACACCAAAAGAGTTTGTATATTAATTCACAAGCTTGATCAAGA	722	
Db	606	665	
0Y	723	GCATGTAAAGAAAGTAGACGGGAAATCATTTATCTACTGTGACAAATATTAATGC	782	
Db	666	725	
0Y	783	TATATCCGGTATTGAAGGACTGTATAAAGGTGATTAATAAGTTTAAACAGAGATAAGA	842	
Db	726	785	
0Y	843	TACC 846		
Db	786	wygc 789		

KW gene expression; ss.
XX Synthetic.
OS WO200107665-A2.
PN 01-FEB-2001.
PD 26-JUL-2000; 2000MO-US20476.
PE 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA Umek RM;
PI WPI; 2001-159728/16.
DR Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses in
XX a single surface -
XX Example 6; Page 128; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 19.1%; Score 162.4; DB 22; Length 936;
Best Local Similarity 0.5%; Pred. No. 9.7e-27;
Matches 4; Conservative 522; Mismatches 258; Indels 0; Gaps 0;

OY 63 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 122
DB 6 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 65
OY 123 GGGTCAAGAGTTATTACACAGCGTGCACATTAATTCATTAAGTAAAGTAAAA 182
DB 66 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 125
OY 183 ATTAGTGAATGTTTCGGGATTTGCAAAATTTGAGTACAGGAAGAAATTTAAGCAAC 242
DB 126 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 185
OY 243 AATATCTGAGATGACAAAGTACGATGATACATCTGGGAATTAATCAAGATGTAC 302
DB 186 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 245
OY 303 GGTTCATTAAGAGTGAAGCGGAAACAAGTGTCTTCTATTAAACCGGACATATGCT 362
DB 246 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 305
OY 363 ACCAGAGATGACGACATGATGATGTTTAAATATTACAAAGAAAAATATATG 422
DB 306 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 365
OY 423 ATGGAAGATATTACTATTAAGAGTCAAGTTCAGTGGACAGTGTGATTTAAGCAC 482
DB 366 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 425
OY 483 ATTAACATTAAATGTCAGAGTACATGACATTAATTAAGTGAACAAAGTCAATAC 542
DB 426 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 485
OY 543 TGAATTTGAAAAAGCCTTTCAGGTTCTAAATTAAGTGTGATTAATACGAAGACCAAT 602
DB 543 TGAATTTGAAAAAGCCTTTCAGGTTCTAAATTAAGTGTGATTAATACGAAGACCAAT 602

DB 486 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 545
OY 603 TGAATTAACATTCACAGCGTATGTCATATATAGTATTTCAATTAACTAAAC 662
DB 546 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 605
OY 663 CAAATTTACGAATGAACAGCAAAAGAGTTTGTAAATTAATCAAGTGTATCAAGA 722
DB 606 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 665
OY 723 GCATGTAAGGAAGGTAACGGGAAATCATTTAATCACTGTCACAAATTTATGC 782
DB 666 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 725
OY 783 TATGCCGATATGAGTACTGTAAAGTGAATTAAGTGTAAACAGATTAAGA 842
DB 726 TACAGTAAGATAGAGGTTATAGTAAACAGTACATTAACTGTTAAAGTGCACAGT 785
OY 843 TACC 846
DB 786 Wwgc 789

RESULT 15

AAF58262
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 19.1%; Score 162.4; DB 22; Length 936;
Best Local Similarity 0.5%; Pred. No. 9.7e-27;
Matches 4; Conservative 522; Mismatches 258; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 21:27:31 ; Search time 90.75 Seconds
(without alignments)
218.786 Million cell updates/sec

Title: US-09-813-820-3

Sequence: 1 GACGATTAATAATGCAAAAT.....AACAGATTAATACCAAG 849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents.NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCBUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	849	100.0	849	4	US-08-856-253-3	Sequence 3, Appl1
2	849	100.0	1500	4	US-08-856-253-5	Sequence 5, Appl1
3	849	100.0	3827	2	US-08-447-031A-1	Sequence 1, Appl1
4	792.2	93.3	4612	2	US-08-447-031A-8	Sequence 8, Appl1
5	441	51.9	441	4	US-08-856-253-1	Sequence 1, Appl1
6	64.2	7.6	7218	1	US-08-232-463-14	Sequence 14, Appl1
7	47.4	5.6	19124	2	US-08-487-826B-13	Sequence 13, Appl1
8	46	5.4	3095	6	5231168-1	Patent No. 5231168
9	42.8	5.0	1341	4	US-08-960-780-51	Sequence 51, Appl1
10	42.8	5.0	1341	4	US-09-073-898-51	Sequence 51, Appl1
11	42.8	5.0	1341	4	US-09-371-913A-2	Sequence 2, Appl1
12	42.8	5.0	4041	1	US-08-471-033-22	Sequence 22, Appl1
13	42.8	5.0	4041	1	US-08-471-044-19	Sequence 22, Appl1
14	42.8	5.0	4041	2	US-08-463-483A-22	Sequence 22, Appl1
15	42.8	5.0	4041	2	US-08-471-046A-22	Sequence 22, Appl1
16	42.8	5.0	4041	2	US-08-470-566B-22	Sequence 22, Appl1
17	42.8	5.0	4041	2	US-08-469-334-22	Sequence 22, Appl1
18	42.8	5.0	4041	3	US-09-300-529-22	Sequence 22, Appl1
19	42.8	5.0	4074	1	US-08-471-033-19	Sequence 19, Appl1
20	42.8	5.0	4074	1	US-08-471-044-19	Sequence 19, Appl1
21	42.8	5.0	4074	2	US-08-463-483A-19	Sequence 19, Appl1
22	42.8	5.0	4074	2	US-08-471-046A-19	Sequence 19, Appl1
23	42.8	5.0	4074	2	US-08-470-566B-19	Sequence 19, Appl1
24	42.8	5.0	4074	2	US-08-469-334-19	Sequence 19, Appl1
25	42.8	5.0	4074	3	US-09-300-529-19	Sequence 19, Appl1
26	42.8	5.0	6049	1	US-08-471-033-1	Sequence 1, Appl1
27	42.8	5.0	6049	2	US-08-471-044-1	Sequence 1, Appl1

28	42.8	5.0	6049	2	US-08-463-483A-1	Sequence 1, Appl1
29	42.8	5.0	6049	2	US-08-471-046A-1	Sequence 1, Appl1
30	42.8	5.0	6049	2	US-08-470-566B-1	Sequence 1, Appl1
31	42.8	5.0	6049	2	US-08-469-334-1	Sequence 1, Appl1
32	42.8	5.0	6049	3	US-09-300-529-1	Sequence 595, App
33	42.4	5.0	658	4	US-08-998-416-595	Sequence 41, Appl
34	41.6	4.9	1689	1	US-07-991-867B-41	Sequence 41, Appl
35	41.6	4.9	1689	2	US-08-544-332-41	Sequence 41, Appl
36	41.6	4.9	8457	1	US-07-991-867B-1	Sequence 1, Appl1
37	41.6	4.9	8457	2	US-08-544-332-1	Sequence 1, Appl1
38	40.8	4.8	615	4	US-08-998-416-186	Sequence 186, App
39	40.8	4.8	5361	4	US-08-973-462-2	Sequence 2, Appl1
40	40.8	4.8	6152	4	US-08-973-462-1	Sequence 1, Appl1
41	40.4	4.8	6768	1	US-08-107-755A-1	Sequence 1, Appl1
42	40.2	4.7	3393	4	US-09-104-324B-1	Sequence 1, Appl1
43	40.2	4.7	3393	4	US-09-162-713-1	Sequence 1, Appl1
44	39.6	4.7	636	4	US-08-998-416-1137	Sequence 1137, Ap
45	39.6	4.7	1511	1	US-07-991-867B-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-856-253-3
Sequence 3, Application US/08856253
Patent No. 6286214
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Patti, Joseph M.
APPLICANT: House-Pompeo, Karen
APPLICANT: Stahanam, Narayana
APPLICANT: Symersky, Jindrich
TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,253
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TAMK:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 849 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-856-253-3
Query Match 100.0%; Score 849; DB 4; Length 849;
Best Local Similarity 100.0%; Pred. No. 4,5e-194;

Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GACGATAAAAATGAAAAATACAAAATGGTGACATGATTAAGTGGCATGGCCGACAAAGC 60
    |||
Db 1 GACGATAAAAATGAAAAATACAAAATGGTGACATGATTAAGTGGCATGGCCGACAAAGC 60
QY 61 GGTACAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 120
    |||
Db 61 GGTACAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 120
QY 121 GTGGGTCAAGCAGTATTAACACGACGCGTCAACAATTAATTCATGATTAAGTAGAA 180
    |||
Db 121 GTGGGTCAAGCAGTATTAACACGACGCGTCAACAATTAATTCATGATTAAGTAGAA 180
QY 181 AATTAAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 240
    |||
Db 181 AATTAAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 240
QY 241 ACAAAATCTTCAGATGACAAAGTACGATTAACATCTGGGAAATTAACAGATGTT 300
    |||
Db 241 ACAAAATCTTCAGATGACAAAGTACGATTAACATCTGGGAAATTAACAGATGTT 300
QY 301 ACGGTTTCATTAAGTGAACGCGGGAACAAGTATGTTTCTATTAATTAACGCGAGATATG 360
    |||
Db 301 ACGGTTTCATTAAGTGAACGCGGGAACAAGTATGTTTCTATTAATTAACGCGAGATATG 360
QY 361 CTACCAAGAAATGACACATGATGATGTTTAAATTAATTAACATGAAGAAATGAT 420
    |||
Db 361 CTACCAAGAAATGACACATGATGATGTTTAAATTAATTAACATGAAGAAATGAT 420
QY 421 GTATCGAAAGATTTACTATTAAGATGATGATGATGATGATGATGATGATGATGAT 480
    |||
Db 421 GTATCGAAAGATTTACTATTAAGATGATGATGATGATGATGATGATGATGATGAT 480
QY 481 ACATTAACATTAATGATGACAGTACACATGATGATGATGATGATGATGATGATGAT 540
    |||
Db 481 ACATTAACATTAATGATGACAGTACACATGATGATGATGATGATGATGATGATGAT 540
QY 541 ACTGATTTTGAAGAAAGCCTTCCAGGTTCTAAATTAATGATGATGATGATGATGATGAT 600
    |||
Db 541 ACTGATTTTGAAGAAAGCCTTCCAGGTTCTAAATTAATGATGATGATGATGATGATGAT 600
QY 601 ATTGATGATTAACATTCACAGGCTATGCTATTAATTAATTAATTAATTAATTAATTA 660
    |||
Db 601 ATTGATGATTAACATTCACAGGCTATGCTATTAATTAATTAATTAATTAATTAATTA 660
QY 661 ACCAAATTAAGTAAAGTAAAGGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 720
    |||
Db 661 ACCAAATTAAGTAAAGTAAAGGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 720
QY 721 GAGCATGTAAGGAAAGTGAACGGGAATCAATTAATCAATCAATCAATCAATCAATTAAT 780
    |||
Db 721 GAGCATGTAAGGAAAGTGAACGGGAATCAATTAATCAATCAATCAATCAATCAATTAAT 780
QY 781 GCTAATGCGGATTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
    |||
Db 781 GCTAATGCGGATTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 GATACCAAG 849
Db 841 GATACCAAG 849

```

TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
 NUMBER OF INVENTION: 8
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: U.S.
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/856,253
 FILING DATE: Concurrently Herewith
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/017,678
 FILING DATE: 16-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: TANK:193
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (512) 474-7577
 INFORMATION FOR SEQ. ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1500 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-856-253-5

Query Match 100.0%; Score 849; DB 4; Length 1500;
 Best local Similarity 100.0%; Pred. No. 5.2e-194;
 Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GACGATAAAAATGAAAAATACAAAATGGTGACATGATTAAGTGGCATGGCCGACAAAGC 60
    |||
Db 94 GACGATAAAAATGAAAAATACAAAATGGTGACATGATTAAGTGGCATGGCCGACAAAGC 153
QY 61 GGTACAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 120
    |||
Db 61 GGTACAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 120
QY 121 GTGGGTCAAGCAGTATTAACACGACGCGTCAACAATTAATTCATGATTAAGTAGAA 180
    |||
Db 121 GTGGGTCAAGCAGTATTAACACGACGCGTCAACAATTAATTCATGATTAAGTAGAA 180
QY 181 AATTAAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 240
    |||
Db 181 AATTAAGTAAAGTAAAGGTTTATGTAATAACGTCATTAACGTGTAAGGTGAACAG 240
QY 241 ACAAAATCTTCAGATGACAAAGTACGATTAACATCTGGGAAATTAACAGATGTT 300
    |||
Db 241 ACAAAATCTTCAGATGACAAAGTACGATTAACATCTGGGAAATTAACAGATGTT 300
QY 301 ACGGTTTCATTAAGTGAACGCGGGAACAAGTATGTTTCTATTAATTAACGCGAGATATG 360
    |||
Db 301 ACGGTTTCATTAAGTGAACGCGGGAACAAGTATGTTTCTATTAATTAACGCGAGATATG 360
QY 361 CTACCAAGAAATGACACATGATGATGTTTAAATTAATTAACATGAAGAAATGAT 420
    |||
Db 361 CTACCAAGAAATGACACATGATGATGTTTAAATTAATTAACATGAAGAAATGAT 420
QY 421 GTATCGAAAGATTTACTATTAAGATGATGATGATGATGATGATGATGATGATGAT 480
    |||
Db 421 GTATCGAAAGATTTACTATTAAGATGATGATGATGATGATGATGATGATGATGAT 480
QY 514 GATACCAAG 514
Db 514 GATACCAAG 514

```


Qy	481	ACATTAAACATTAAATGTCACAGGTACACATAGCAATATTATATAGTGACACAAAGTCAATT	540
Db	574	ACATTAAACATTAAATGTGACAGGTACACATAGCAATATTATATAGTGACACAAAGTCAATT	633
Qy	541	ACGTATTTTGAAGAAAGCCTTCCAGTTCGTTCAATTAATCTGTGATATACGAAGACACA	600
Db	634	ACTGATTTTGAAGAAAGCCTTCCAGTTCCTAAATTAATCTGTGATATACGAAGACACA	693
Qy	601	ATTGATGTAAACAATTCACAGAGCTATGGTCAATATAATAGTTTTCATTAACACTACAAA	660
Db	694	ATTGATGTAAACAATTCACAGAGCTATGGTCAATATAATAGTTTTCATTAACACTACAAA	753
Qy	661	ACCAAAATTACGAATGGAACAGCAAAAAAGATTGTATATATTCACAGCTTGGTATCAA	720
Db	754	ACCAAAATTACGAATGGAACAGCAAAAAAGATTGTATATATTCACAGCTTGGTATCAA	813
Qy	721	GAGCATGTGAAGAAAGTGAACAGGAAATCATTTAATCATCTGTGCACATATATTAAT	780
Db	814	GAGCATGTGAAGAAAGTGAACAGGAAATCATTTAATCATCTGTGCACATATATTAAT	873
Qy	781	GCTAATGCCGTTATGGAAGGTACTGTAAAAAGTGCAATTAAAGTTTAAAAACGATATAA	840
Db	874	GCTAATGCCGTTATGGAAGGTACTGTAAAAAGTGCAATTAAAGTTTAAAAACGATATAA	933
Qy	841	GATACCAAG 849	
Db	934	GATACCAAG 942	

RESULT 3
US-08-447-031A-1
Sequence 1, Application US/08447031A
Patent No. 5851794
GENERAL INFORMATION:
APPLICANT: GUSS, Bengt
APPLICANT: HOOK, Magnus
APPLICANT: JONSSON, Hans
APPLICANT: LINDBERG, Martin
APPLICANT: PATTI, Joseph
APPLICANT: SIGMANS, Christer
APPLICANT: SWITLASKI, Lech
TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
TITLE OF INVENTION: ITS PREPARATION
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P. O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447, 031A
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/861, 804
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/SE91/00707
FILING DATE: 22-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9003374-7
FILING DATE: 22-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300

```

REFERENCE/DOCKET NUMBER: 012869-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3827 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-0447-031A-1

```

Query Match	100.0%;	Score 849;	DB 2;	length 3827;
Best Local Similarity	100.0%;	Pred. No. 6.6e-194;		
Matches 849;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	GACGATATAAAATGGAAAAATATCAAAATGTGGACATCATTTAAAGTGGCATGGCCGCAACG	60
Db	331	GACGATATAAAATGGAAAAATATCAAAATGTGGACATCATTTAAAGTGGCATGGCCGCAACG	390
QY	61	GGTACAGTAAAAAGATAGAGGGTTATAGTAAAAACGTACATTAATCTGTTTAAAGGTAAACAG	120
Db	391	GGTACAGTAAAAAGATAGAGGGTTATAGTAAAAACGTACATTAATCTGTTTAAAGGTAAACAG	450
QY	121	GTGGGTCAAGCAGTTATTTACACGACGAGCGTGCACAACTTACATTCATATTAAGTAGAA	180
Db	451	GTGGGTCAAGCAGTTATTTACACGACGAGCGTGCACAACTTACATTCATATTAAGTAGAA	510
QY	181	AAATTAAGTGAATGTTTGGGATTTTGCAGATTTTGAAGTACAAAGAAATTTAACCGAA	240
Db	511	AAATTAAGTGAATGTTTGGGATTTTGCAGATTTTGAAGTACAAAGAAATTTAACCGAA	570
QY	241	ACAAATACTTCAGATGACAAAGTAGTACGATTAACATCTGGGAATTAATCAACGAATGT	300
Db	571	ACAAATACTTCAGATGACAAAGTAGTACGATTAACATCTGGGAATTAATCAACGAATGT	630
QY	301	ACGGTTCATTAAGTGAAGCGGGACACAGTAGTGTGTTTCTATTTATTAACCGGAGATATG	360
Db	631	ACGGTTCATTAAGTGAAGCGGGACACAGTAGTGTGTTTCTATTTATTAACCGGAGATATG	690
QY	361	CTACACAGAGTATACGACACATGTACGATGGTTTTTAAATTAACAAAGAAAAAGTAT	420
Db	691	CTACACAGAGTATACGACACATGTACGATGGTTTTTAAATTAACAAAGAAAAAGTAT	750
QY	421	GTATCGAAAGATATTTACTATTAAGAGATCAGATTCAGAGTGGACAGCAGTATGATTTAAGC	480
Db	751	GTATCGAAAGATATTTACTATTAAGAGATCAGATTCAGAGTGGACAGCAGTATGATTTAAGC	810
QY	481	ACATTTAAACATTTAATGTGACAGGTACACATATGATTTATATATGTGGACAAAGTCAATT	540
Db	811	ACATTTAAACATTTAATGTGACAGGTACACATATGATTTATATATGTGGACAAAGTCAATT	870
QY	541	ACTGATTTTGGAAAAAGCTTCCAGGTTCTTAAATAAGCTTGATTAATACGAAGAACACA	600
Db	871	ACTGATTTTGGAAAAAGCTTCCAGGTTCTTAAATAAGCTTGATTAATACGAAGAACACA	930
QY	601	ATTGATGTAAACATTTCCACAAGGCTATGGGTCTATATATAGTTTTTCAATTAACATCAAA	660
Db	931	ATTGATGTAAACATTTCCACAAGGCTATGGGTCTATATATAGTTTTTCAATTAACATCAAA	990
QY	661	ACCAAAATTTACGAATGAACAGCAAAAAAGATTTGTTAATTAATTCACAGCTTGGTATCAA	720
Db	991	ACCAAAATTTACGAATGAACAGCAAAAAAGATTTGTTAATTAATTCACAGCTTGGTATCAA	1050
QY	721	GAGCATGGTAAAGAAAGTGAACCGGAAATCTTTAATCATACATGTCGACACAATTTATAT	780
Db	1051	GAGCATGGTAAAGAAAGTGAACCGGAAATCTTTAATCATACATGTCGACACAATTTATAT	1110
QY	781	GCTAATGCCGCTATTGAAGGTACTGTAAAGGTGAATTTAAAGTTTTTAAACAGGATATAA	840
Db	1111	GCTAATGCCGCTATTGAAGGTACTGTAAAGGTGAATTTAAAGTTTTTAAACAGGATATAA	1170

```

Oy      841 GATACCAAG 849
|||||||
Db      1171 GATACCAAG 1179

RESULT      4
US-08-447-031A-8
; Sequence 8, Application US/08447031A
; Patent No. 5851794
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: HOOK, Magnus
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: PATTI, Joseph
; APPLICANT: SIGNAS, Christer
; APPLICANT: SWITALSKI, Lech
; TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
; TITLE OF INVENTION: ITS PREPARATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,031A
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/861,804
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE91/00707
; FILING DATE: 22-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 900374-7
; FILING DATE: 22-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 012889-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ. ID NO. 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4612 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 931..4485
;
US-08-447-031A-8

Query Match      93.3%; Score 792.2; DB 2; Length 4612;
Best Local Similarity 97.8%; Pred. No. 2.4e-180;
Matches 836; Conservative 0; Mismatches 13; Indels 6; Gaps

```

QY	61	GGTCACTAAAGATAGAGGGTTTATGATAAAACAGTACATTAACGTGTTAAAGGTAAACAG	120
Db	1171	GGTCACTAAAGATAGAGGGTTTATGATAAAACAGTACATTAACGTGTTAAAGGTAAACAG	1230
QY	121	GTGGGTGACAGCATTTTATTAACACAGACGGTGCACAACTTACATTCATGATTAAGTAGAA	180
Db	1231	GTGGGTGACAGCATTTTATTAACACAGACGGTGCACAACTTACATTCATGATTAAGTAGAA	1290
QY	181	AAATTAAGTATGTTTCGGGATTTGCAGAAATTTGAAGTACAAAGGAATTTTAACGGA	240
Db	1291	AAATTAAGTATGTTTCGGGATTTGCAGAAATTTGAAGTACAAAGGAATTTTAACGGA	1350
QY	241	ACAATATCTTCAGTGCAGAAAGTAGCTAGATTAACATCTGGGAATAATATCAACGAATGTT	300
Db	1351	ACAATATCTTCAGTGCAGAAAGTAGCTAGATTAACATCTGGGAATAATATCAACGAATGTT	1410
QY	301	A - -CGGTCATTAAGAAAGTGAAGCGGGAACAAGTAGTGTGTTCTATTATATATAA - -CGGAG	355
Db	1411	ATCGGTTGATTAAGAAAGTGAAGCGGGAACAAGTAGTGTGTTCTATTATATATATAA - -CGGAG	1470
QY	356	ATATGCTAAC - AAGAGTAGACAGACATGTACATGTTTAAATATTAACATTAAGAAAA	414
Db	1471	ATATGCTAACAGAAAGTATACACATGTACATGTTTAAATATTAACATTAAGAAAA	1530
QY	415	AGTATGTATCGAAAGATATTACTATTAAAGATCAGATTCAGAGTGCAGCAGCTAGAT	474
Db	1531	AGTATGTATCGAAAGATATTACTATTAAAGATCAGATTCAGAGTGCAGCAGCTAGAT	1590
QY	475	TTAAGCAGATTAACATTTAATGTGACAGGTACACATAGCAATTTATTATGTGACAAAGT	534
Db	1591	TTAAGCAGATTAACATTTAATGTGACAGGTACACATAGCAATTTATTATGTGACAAAGT	1650
QY	535	GCAATTCAGTATTTTGAAAAAGCCTTCCAGGTTCCAAATTAACGTTGATATACGGAAG	594
Db	1651	GCAATTCAGTATTTTGAAAAAGCCTTCCAGGTTCCAAATTAACGTTGATATATACGGAAG	1710
QY	595	AACCAATGATGTAACAAATTCACAAAGCTATGAGTCAATATAAGTTTTCATTTAAC	654
Db	1711	AACCAATGATGTAACAAATTCACAAAGCTATGAGTCAATATAAGTTTTCATTTAAC	1770
QY	655	TACAAAACCAAAATTACGAATGAACAGCAAAAAGATTGTTAATTAATTCACACAGCTGG	714
Db	1771	TACAAAACCAAAATTACGAATGAACAGCAAAAAGATTGTTAATTAATTCACACAGCTGG	1830
QY	715	TATCAAGAGATGCTAGGAAGAAAGTGAACGGGAATCAATTAATCATCTGTGCAAT	774
Db	1831	TATCAAGAGATGCTAGGAAGAAAGTGAACGGGAATCAATTAATCATCTGTGCAAT	1890
QY	775	ATTAAATCTATGCCGCTATTGGAAGTACTGTAAAGGTGAATTAAGTTTAAAAACAG	834
Db	1891	ATTAAATCTATGCCGCTATTGGAAGTACTGTAAAGGTGAATTAAGTTTAAAAACAG	1950
QY	835	GATTAAGATACCAAG	849
Db	1951	GATTAAGATACCAAG	1965
RESULT 5			
US-08-856-253-1			
; Sequence 1, Application US/08856253			
; Patent No. 6288214			
; GENERAL INFORMATION:			
; APPLICANT: Hook, Magnus			
; APPLICANT: Patti, Joseph M.			
; APPLICANT: House-Pompeo, Karen			
; APPLICANT: Stranam, Nairajana			
; APPLICANT: Symersky, Jindrich			
; TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS			
; NUMBER OF SEQUENCES: 8			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Arnold, White & Durkee			

```

STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,253
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TANK:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-856-253-1

```

```

Query Match      51.9%; Score 441; DB 4; Length 441;
Best Local Similarity 100.0%; Pred. No. 7.6e-97;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 271 ATACATCTGGGATAATCAAGATGTACGCTCATAAAGTAAAGCGGAGAACAGT 330
DB 1 ATACATCTGGGATAATCAAGATGTACGCTCATAAAGTAAAGCGGAGAACAGT 60
QY 331 AGTGTCTTATATATAAAGCGGAGATGCTACAGAAATGACAGATGACATGACATG 390
DB 61 AGTGTCTTATATATAAAGCGGAGATGCTACAGAAATGACAGATGACATGACATG 120
QY 391 TTTTAAATATTAACAATGAAAAAGTATGATCGAAAAATATTAAGATCAG 450
DB 121 TTTTAAATATTAACAATGAAAAAGTATGATCGAAAAATATTAAGATCAG 180
QY 451 ATTCAAGTGGACAGAGTATGATTAACACATTAACATTAATGACAGTACACAT 510
DB 181 ATTCAAGTGGACAGAGTATGATTAACACATTAACATTAATGACAGTACACAT 240
QY 511 ACAATATATATAGTGAACAAGTCAATTAATGATTTGAAAAAGCTTCCAGTTCT 570
DB 241 ACAATATATATAGTGAACAAGTCAATTAATGATTTGAAAAAGCTTCCAGTTCT 300
QY 571 AAAATTAATGTTGATTAATGAGAAACACAAATGATTAACATTCACAAAGCTATGG 630
DB 301 AAAATTAATGTTGATTAATGAGAAACACAAATGATTAACATTCACAAAGCTATGG 360
QY 631 TCATATATATAGTTTTCATTAATCAAAACCAAAATTAACGATTAACAGCAAAAGAG 690
DB 361 TCATATATATAGTTTTCATTAATCAAAACCAAAATTAACGATTAACAGCAAAAGAG 420
QY 691 TTTGTTAATTAATTCACAGCT 711
DB 421 TTTGTTAATTAATTCACAGCT 441

```

```

RESULT 6
US-08-232-463-14/c
Sequence 14, Application US/08232463

```

```

Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
CLONE: PT9pct-F1s
US-08-232-463-14

```

```

Query Match      7.6%; Score 64.2; DB 1; Length 7218;
Best Local Similarity 6.5%; Pred. NO. 1.1e-06;
Matches 27; Conservative 225; Mismatches 163; Indels 0; Gaps 0;

```

```

QY 101 TAACTGTTAAGGTGAAGGTGGTCAAGCACTTATTAACACGACGCGCAACATTA 160
DB 1474 TATCTATGCAAGTAGTTAAAGACATAGAACAAATTGTCACRRRRRRRRRRRRRR 1415
QY 161 CATTCATGATAAGTAAGAAAAATTAAGTATGTTCCGGGATTTGACAAATTTGAAGTAC 220
DB 1414 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1355
QY 221 AAGGAAGAATTTAAGCAACAATACTTCAGATACACAAAGTAGTACATTAACATCTG 280
DB 1354 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1295
QY 281 GGAATAATCAACAGATGTTACGTTCAATTAAGTGAAGCGGAGAACAGTAAGTTTCT 340
DB 1294 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1235
QY 341 ATTATAAAGCGAGATATGCTACAGAAAGATACGACACATGTACAGTGTTTTAAATA 400
DB 1234 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1175
QY 401 TTACATCAAAAAAGTTATGATCAAAAGATATTCTATAAGATCAATTCAGATCAAGCTG 460

```


Db 1785 agtagaggtgaagaatctccagaatgttgtaaatgaagaagtaccacaaac 1844
Qy 576 AACGTGTATATACGAGACACATGATTAACATTCACAGGCTTGGTCTATA 635
Db 1845 aaatacaatgaactatgaactataaacccagaagaaaaagaatgaattagtg 1904
Qy 636 TAATAGTTTTCAATTAATCAACAAACCAAAATTAACGAATGACGAAAGAGTTGT 695
Db 1905 tgaagaaaagaacatccacagaagaccggtgtacctatcaatgaataatgaataacgt 1964
Qy 696 TAATTAATTCACAGCTGTGATCAAGACCATGTAAGGAGAGAGTGAACGGAAATCATT 755
Db 1965 tactcccaacacatctgaggtgactccactaaaccagataatagttcaaatlaaatag 2024
Qy 756 TAATCATCTGTGCACATATTAATGCTAATGCCGGTTTGAAGTACTGTAAGAGTGA 815
Db 2025 acaagaaataaaccacaaataaagaaacacagtagtagtggtccaaaacatgtag 2084
Qy 816 ATTAAGTTTAAACAGATTAAGAT 843
Db 2085 acaaatatacagaagatgataatgat 2112

RESULT 9
US-08-960-780-51
; Sequence 51, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Felleison, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 177C8
; US-08-960-780-51

Query Match 5.0%; Score 42.8; DB 4; Length 1341;
Best Local Similarity 47.4%; Pred. No. 0.091;
Matches 128; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Qy 393 TTTAAATATTAACATGAAAAAGTTATGATTCGAAAGATATTACTATTAAGATCAGAT 452
Db 288 TATAAGACAAATTTATAAGAAATTAATCTTTCTATGCGACGCTCATTTGAAGATGAAT 347
Qy 453 TCAAGCTGCAGACAGCTAGATTATAGACATTAACATTTAATGATGACAGATACATAG 512
Db 348 AAAAGATTTAAAGAAATTTAGATTAAGATTTGATTAACCAATCTATCAAAATTTCTATTAT 407
Qy 513 CAATTATATATGTCGACAAAGTCGATTAATTTGAAAAAGCCTTCCAGTTCTAA 572
Db 408 CACCTATATAAATGTGACACCGACACAAATGGATTTAATTAATCTTTAACAGAAAGTAA 467
Qy 573 AATACTGTGATATATACGAAAGAACACAAATGATGTACAAATTCACAAAGGCTATGGGTC 632
Db 468 TACGATTAATTCGTGATGCAATGCGACACTTTAAAGAACAAATTTAGATAGGATATTAA 527
Qy 633 ATATATATGTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 662
Db 528 GTTGATGATTTATCTAGATGACGATTTTAAAC 557

RESULT 10
US-09-073-898-51
; Sequence 51, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Felleison, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Finstad-Lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.

THIS PAGE BLANK (USPTO)